

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 05:45:20 ; Search time 3396 Seconds  
(without alignments)  
10872.460 Million cell updates/sec

Title: US-09-441-857-11

Perfect score: 762

Sequence: 1 atgtccggcgaggagactt.....tgctgcctgtctcgtggaga 762

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_com.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	761	99.9	765	9	CR533469
2	761	99.9	840	9	AF309646 Homo sapi
3	761	99.9	3084	6	AX086357 Sequence
4	759.4	99.7	198961	9	AC079776 Homo sapi
5	759.4	99.7	349981	6	AX647880
6	759.4	99.7	349989	6	AX647177
7	757.8	99.4	165411	9	AC018865
8	755.6	99.2	195280	9	AC073869 Homo sapi
9	744.4	97.7	242865	2	AC120839
10	741.2	97.3	184706	9	AC120781
11	684	89.8	3070	9	BC003617
12	684	89.8	3100	9	BC068486
13	684	89.8	3120	9	AK057157
14	639.2	83.9	3745	6	AR447780
15	638.2	83.8	1770	10	BC019118
16	609.2	79.9	739	9	AF119836
17	591.8	77.7	683	9	AF130122
18	587.6	77.1	627	9	AF498941
19	586	76.9	627	9	AF198616

20	586	76.9	627	9	CR541984
21	585.4	76.8	1950	10	AB041575
22	559.6	73.4	740	6	AX771444
23	559.6	73.4	740	9	HUMRAB6A
24	554.8	72.8	1170	5	CR390639
25	542.8	71.2	627	9	AF130986
26	542.8	71.2	627	9	AF498939
27	532.6	69.9	844	5	AF540409
28	526.2	69.1	2527	10	BC026915
29	526	69.0	573	9	AF124200
30	477.8	62.7	172862	10	AL713865
31	471.6	61.9	2099	5	BC046683
32	429	56.3	1461	5	BC064230
33	401	52.6	2456	6	AX713746
34	401	52.6	2456	9	AK055504
35	384.8	50.5	1261	9	BC002510
36	384.4	50.4	627	9	BT007263
37	384.4	50.4	2498	5	BC058059
38	383.2	50.3	1266	9	AF166492
39	383.2	50.3	1589	6	BD131169
40	382.8	50.2	627	9	AF498940
41	381.6	50.1	2005	5	BC080215
42	381.6	50.1	5395	9	BC078662
43	375.4	49.3	4768	10	BC060618
44	369	48.4	1652	5	BC074238
45	352.2	46.2	3826	6	AX647191

## ALIGNMENTS

CR533469 765 bp mRNA linear PRI 22-JUN-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834C0517D for gene RAB6C, RAB6C, member RAS oncogene family; complete cds, incl. stopcodon.

CR533469 GI:49065363

Full ORF shuttle clone, Gateway(TM), complete cds.

Homo sapiens (human)

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR533469 RZPD0834C0517D, ORFNO 2784  
RZPD; RZPD0834C0517D, ORFNO 2784  
www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPD0834C0517D RZPDLIB;  
Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.  
834  
www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834  
www.rzpd.de/products/orfclones/  
Contact: Inge Arlart  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 100  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available from RZPD;  
contact RZPD (customer.service@rzpd.de) for further information.  
This CDS clone is a part of a collection of human full ORF clones  
generated by RZPD.  
This CDS has been cloned incl. stopcodon.  
The CDS has been inserted into pDONR201 via a BP Clonase(TM)













## AUTHORS

Kim, K., Haglund, K., Tomlinson, C., Mangiapanello, L. and Dignan, G.  
The sequence of Homo sapiens BAC clone RP11-32C20  
Unpublished (2001)  
3 (bases 1 to 165411)

## REFERENCE

## JOURNAL

Waterston, R.H.  
Direct Submission  
Submitted (21-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 165411)

## REFERENCE

## AUTHORS

## JOURNAL

Waterston, R.H.  
Direct Submission  
Submitted (13-SEP-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 165411)

## REFERENCE

## AUTHORS

## JOURNAL

Waterston, R.  
Direct Submission  
Submitted (23-NOV-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 23, 2002 this sequence version replaced gi:22830544.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_NH0032C20  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoh, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

There is an unresolved tandem repeat from base 115267 to 124390.

Polymorphisms have been identified between AC018804 and this sequence.

Data from AC018804 was used to finish this clone.

This sequence is the entire insert of the clone.

Location/Qualifiers

1..165411

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

## FEATURES

## source

```

/chromosome="2"
/map="2"
/clone="RP11-32C20"
/clone_lib="RPCI-11"
487..1684
/rpt_family="L1"
repeat_region
1678..1885
/rpt_family="L1"
repeat_region
1890..1986
/rpt_family="L1"
repeat_region
1987..2289
/rpt_family="Alu"
repeat_region
2294..2380
/rpt_family="GA-rich"
repeat_region
2392..2699
/rpt_family="L1"
repeat_region
2905..3262
/rpt_family="Alu"
repeat_region
3284..3395
/rpt_family="T-rich"
repeat_region
3601..3674
/rpt_family="MER105"
repeat_region
3726..3845
/rpt_family="Alu"
repeat_region
3846..3869
/rpt_family="(TAA)n"
repeat_region
3870..4042
/rpt_family="Alu"
repeat_region
4064..4364
/rpt_family="ERV1"
repeat_region
4692..4980
/rpt_family="Alu"
repeat_region
4981..5001
/rpt_family="AT-rich"
repeat_region
5006..5316
/rpt_family="Alu"
repeat_region
5340..5519
/rpt_family="ERV1"
repeat_region
5563..5604
/rpt_family="AT-rich"
repeat_region
5637..5678
/rpt_family="AT-rich"
repeat_region
5757..5781
/rpt_family="AT-rich"
repeat_region
5782..6005
/rpt_family="Alu"
repeat_region
6067..6146
/rpt_family="MER1_type"
repeat_region
6196..6487
/rpt_family="Alu"
repeat_region
6515..6634
/rpt_family="MER1_type"
repeat_region
6755..7044
/rpt_family="Alu"
repeat_region
7349..7641
/rpt_family="Alu"
repeat_region
7644..7932
/rpt_family="Alu"
repeat_region
7948..8108
/rpt_family="BC200"
repeat_region
8294..8589
/rpt_family="Alu"
repeat_region
8824..9036
/rpt_family="L1"
repeat_region
9037..9342
/rpt_family="Alu"
repeat_region
9343..9404
/rpt_family="L1"
repeat_region
9405..9444
/rpt_family="Alu"
repeat_region
9445..9472
/rpt_family="(TTTTG)n"
repeat_region
9473..9748

```



clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J. and de Jong P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-433A19. Actual start of this clone is at base position 1 of RP11-109E12; actual end is at base position 195280 of RP11-109E12.

Data from AC022629 was used to finish AC073869.

## FEATURES

```
Location/Qualifiers
1. .195280
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
```

misc feature 1. .167/crone

```
1. .167
/crone_11b="RFC1-11"
/note="match to EST AA904412 (NTD:G3030535) 0607e01 91"
```

misc	feature	82	.96
------	---------	----	-----

```

/note="similar to Homo sapiens EST AV651649
82. .96
/note="Maccii CO EST AA504412 (NID:Y303333)

```

misc feature	298	7
(NID:99)	298	7

```
(NID:g9872663)"
298. .778
/notoc"match to EST AW202518 (NID:G570025A)"
```

misc feature	408	4
/hotel="1	408	4

```

/note="match to ESI AW293618 (NID:96700254)
408. .431
/note="similar to HOMO cations EST AV551548

```

misc feature 503 6 (NID:99)

```
(NID:998/2663):
503: .670
/notes="match" C PST AT32094 (NID:24088290) 2710601 y1"
```

misc feature 651 1 /note="

```
651. .1100  
/note=="match to ES1 A1332084 (RID:g4089230) q1r0b0v1.x1"  
EST A1130590 (RID:c3644552) c305h12 v1"
```

misc	feature	695	7

```

/NOTE="MATCH TO ESI A1138380 (NID:G3644332)
695. .710
/NOTE="SIMILAR TO YOMC GANJONG EST BE281748

```

misc feature	(NID:gl 1698
--------------	--------------

(NID:gl0212946)"  
1698. .2014  
/note="similar to Homo sapiens EST N30408 (NID:gl148928)"

WY13C09  
2015

YY13C09.81"  
2015. .2410  
/note="similar to Home 0001.000 ECM AYCE1C40

ENCLOSURE  
(NID: 99-7252)

(NID:99872663)"  
2253. .2732

```

m1.c 304420
m1.c 304420
/note="1"
3330

```

```

/note="match to EST AW340014 (NID:g6836640)
2338. 3119

```

(NID:gl  
2240

```
{NID:g11000861}"
2340. .2885
```

misc\_feature 2340. :  
/not a-"  

2340. .2885  
/note="similar to Bos taurus EST BE722624 (  
2154 2041

misc_feature	/note="similar to Canis familiaris EST BI430421 (NID:g15207653)"
misc_feature	3236..3645
misc_feature	/note="similar to Mus musculus EST A617192 (NID:g2504397)"
misc_feature	vj79a09.x1"
misc_feature	4375..4376
misc_feature	/note="similar to Homo sapiens EST A1677857 (NID:g4888039)"
misc_feature	wd34c04.x1"
repeat_region	4427..4468
repeat_region	/rpt_family=" (TTTA)n"
repeat_region	4440..4738
repeat_region	/rpt_family="Alu"
repeat_region	4835..4973
repeat_region	/note="similar to Homo sapiens EST A1677857 (NID:g4888039)"
repeat_region	wd34c04.x1"
repeat_region	4892..5127
repeat_region	/note="similar to Homo sapiens EST BF511460 (NID:g11594758)"
repeat_region	4899..5106
repeat_region	/note="match to EST AW804021 (NID:g7855891)"
repeat_region	4974..5130
repeat_region	/note="similar to Homo sapiens EST A1677857 (NID:g4888039)"
repeat_region	wd34c04.x1"
repeat_region	5205..5387
repeat_region	/note="similar to Homo sapiens EST BF511460 (NID:g11594758)"
repeat_region	5368..5470
repeat_region	/rpt_family=" (CGG)n"
repeat_region	6103..6166
repeat_region	/rpt_family=" (GGGA)n"
repeat_region	6384..6390
repeat_region	/note="match to EST AL042663 (NID:g5935522)"
repeat_region	6456..6485
repeat_region	/rpt_family=" (CA)n"
repeat_region	6550..6732
repeat_region	/rpt_family="MIR"
repeat_region	7058..7505
repeat_region	/note="similar to Homo sapiens EST BE646288 (NID:g9970599)"
repeat_region	7216..7316
repeat_region	/rpt_family="MIR"
repeat_region	7327..7378
repeat_region	/rpt_family="MERI_type"
repeat_region	8007..8043
repeat_region	/rpt_family="AT_rich"
repeat_region	8009..8118
repeat_region	/rpt_family="Alu"
repeat_region	8123..8336
repeat_region	/rpt_family="L1"
repeat_region	8338..8442
repeat_region	/rpt_family="L1"
repeat_region	8443..8752
repeat_region	/rpt_family="Alu"
repeat_region	8753..9325
repeat_region	/rpt_family="L1"
repeat_region	9488..9910
repeat_region	/note="match to EST BE973696 (NID:g10587032)"
repeat_region	9763..9944
repeat_region	/note="similar to Homo sapiens EST BG739917 (NID:g14050570)"
repeat_region	9763..9944
repeat_region	/note="similar to Homo sapiens EST R18754 (NID:g772364)"
repeat_region	9767..9981
repeat_region	/note="match to EST BE859007 (NID:g10374638)"
repeat_region	10270..10476
repeat_region	/rpt_family="MIR"
repeat_region	10803..10855
repeat_region	/rpt_family="AT_rich"
repeat_region	11485..11542
repeat_region	/note="match to EST BE973696 (NID:g10587032)"
repeat_region	11771..12131
repeat_region	/rpt_family="L2"

misc\_feature 11798..11948  
/note="match to EST BE973696 (NID:g10587032)"  
misc\_feature 12659..12660  
/note="similar to Homo sapiens EST AW891034 (NID:g8055239)"  
misc\_feature 12666..12684

Query Match 99.2%; Score 755.6; DB 9; Length 195280;  
Best Local Similarity 99.5%; Pred. No. 8.5e-220;  
Matches 758; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DQ 37987 ATGTCCCGGGCGGAGACTTCGGGAATCCGCTGAGAAATTCAAGCTGGTGTCTCTGGGG 60  
|||  
Db 37987 ATGTCCCGGGCGGAGACTTCGGGAATCCGCTGAGAAATTCAAGCTGGTGTCTCTGGGG 37928  
|||||  
QY 61 GAGCAAGCCTTGCAAGACATCTTTGATCACAGATTCAGGTATGACAGATTTCACAAAC 120  
|||||  
Db 37927 GAGCAAGCCTTGCAAGACATCTTTGATCACAGATTCAGGTATGACAGATTTCACAAAC 37868  
|||||  
QY 121 ACCTATCAGCAATAAATTTGCATTTGACTTTTATCAAAAATCATGTCTGGAGATGGA 180  
|||||  
Db 37867 ACCTATCAGCAATAAATTTGCATTTGACTTTTATCAAAAATCATGTCTGGAGATGGA 37808  
|||||  
QY 181 ACAATCGGCTTCGGCTGCGGATACGGGGGTACGGAACGCTCTCGTAGCCTCAATCCC 240  
|||||  
Db 37807 ACAATCGGCTTCGGCTGCGGATACGGGGGTACGGAACGCTCTCGTAGCCTCAATCCC 37748  
|||||  
QY 241 AGGTACATCCGTGATCTCTGTCAGCTGTAGTAGTTTACGATATCACAAATGTTAACTCA 300  
|||||  
Db 37747 AGGTACATCCGTGATCTCTGTCAGCTGTAGTAGTTTACGATATCACAAATGTTAACTCA 37688  
|||||  
QY 301 TTCAGCAAACTACAAGTGGATTTGATGTGACAGAACAGAAAGGAGGATGATGTATC 360  
|||||  
Db 37687 TTCAGCAAACTACAAGTGGATTTGATGTGACAGAACAGAAAGGAGGATGATGTATC 37628  
|||||  
QY 361 ATCAGCTAGTAGAATAAGACAGATCTTGTCTGACAGAGCAAGTGTCTAGTTGAGGAG 420  
|||||  
Db 37627 ATCAGCTAGTAGAATAAGACAGATCTTGTCTGACAGAGCAAGTGTCTAGTTGAGGAG 37568  
|||||  
QY 421 GGAGAGGAGAAAGCCAAAGGGCTGATGTAGTTTATTGAACCTAGGCGCAAAACCTGGA 480  
|||||  
Db 37567 GGAGAGGAGAAAGCCAAAGGGCTGATGTAGTTTATTGAACCTAGGCGCAAAAGCTGGA 37508  
|||||  
QY 481 TACAATGTAAAGCAGCTCTTTTCAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGACACA 540  
|||||  
Db 37507 TACAATGTAAAGCAGCTCTTTTCAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGACACA 37448  
|||||  
QY 541 CAGGACGGAAGCAGAGACATGAGTGACATATAAACTGGAAAAACCTCAGAGCAAAACA 600  
|||||  
Db 37447 CAGGACGGAAGCAGAGACATGAGTGACATATAAACTGGAAAAACCTCAGAGCAAAACA 37388  
|||||  
QY 601 GTCAAGGAGAGGGGTTGTTCTCTGCTACTCTCCCATGTATCTTCAACCTCTCTCAGAG 660  
|||||  
Db 37387 GTCAAGGAGAGGGGTTGTTCTCTGCTACTCTCCCATGTATCTTCAACCTCTCTCAGAG 37328  
|||||  
QY 661 CCCCTTACTCTTTCAFTGACTGCAGTGTAATATTGGCTTGAACCTTTTCCCTTCATTA 720  
|||||  
Db 37327 CCCCTTACTCTTTCAFTGACTGCAGTGTAATATTGGCTTGAACCTTTTCCCTTCATTA 37268  
|||||  
QY 721 ATAACGTTTGGCAATTCATCATGCTGCTCTGCTCGTGAGA 762  
|||||  
Db 37267 ATAACGTTTGGCAATTCATCATGCTGCTCTGCTCGTGAGA 37226  
|||||

RESULT 9  
AC120839/c  
LOCUS  
DEFINITION Pan troglodytes clone rp43-5n22, WORKING DRAFT SEQUENCE, 3 ordered pieces.  
ACCESSION AC120839  
VERSION AC120839.38 GI:45120242  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE Pan troglodytes (Chimpanzee)

[illegible]



Db	86875	ATAACGTTTTCGAATTCATCTGCTCCCTGTCCTCATGGAGA	86916	
RESULT 11	BC003617			
LOCUS	BC003617			
DEFINITION	1. mRNA (cDNA clone MGC:1654 IMAGE:3506585), complete cds.			
ACCESSION	BC003617			
VERSION	BC003617.2	GI:33870607		
KEYWORDS	MGC.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 3070) Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uesdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Harte, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalish, D.E., Schneringer, A., Schein, J.E., Jones, S.J., and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 3070) Strausberg, R.			
AUTHORS	Direct Submission			
TITLE	Submitted (26-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
JOURNAL	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> On Aug 19, 2003 this sequence version replaced gi:13177663. Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: DCTD/BTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology <a href="http://www.systemsbio.org">http://www.systemsbio.org</a> contact: <a href="mailto:amadon@systembiology.org">amadon@systembiology.org</a> Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting			
REMARK	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 7 Row: a Column: 8 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 19923230. Location/Qualifiers 1. .3070 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:1654 IMAGE:3506585" /tissue_type="Kidney, renal cell adenocarcinoma" /clone_lib="NIH MGC 14" /lab_host="DH10B-R" /note="vector: pOTB7"			
COMMENT				
FEATURES	source			
gene	1. .3070 /gene="RAB6A" /note="synonym: RAB6" /db_xref="LocusID:5870" /db_xref="MIM:179513" 418..1044 /gene="RAB6A" /codon_start=1 /product="RAB6A, member RAS oncogene family, isoform a" /protein_id="AAH03617.1" /db_xref="GI:13177664" /db_xref="LocusID:5870" /db_xref="MIM:179513" /translation="MSGGDFGNPLRKLKFLVGEQSVGKTSILTRPMYDSFNTYQA TTGIDFLSKTWLYLEDRIRLQWMDTKADLRQVSIIEBERKAKELNVVFIETSAKAG QTTWKIDDDVETERGSDVIMLVNKTDLADLRQVSIIEBERKAKELNVVFIETSAKAG YNVKQLFRVAAALPGMESTQDRSRDMDIKLEKPOEQPVSEGGSC"			
CDS	Query Match 89.8%; Score 684; DB 9; Length 3070; Best Local Similarity 93.9%; Pred. No. 4.4e-198; Matches 727; Conservative 0; Mismatches 35; Indels 12; Gaps 1; ORIGIN 1 ATGTCGCGCGGCGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTTCCTCGGG 60 418 ATGTCACGCGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTTCCTCGGG 477 61 GAGCAAAAGCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTGCAAC 120 478 GAGCAAAAGCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTGCAAC 537 121 ACCTATCAGCAATAATTGGCAATGACTTTTATCAAAAATCATGTAATTTGGAGGATGGA 180 538 ACCTATCAGCAATAATTGGCAATGACTTTTATCAAAAATCATGTAATTTGGAGGATGGA 597 181 ACAATCGGCTTGGCTGTGGGATACCGCGGGTCAGGAACGTCCTCGTAGCTCATTTCCC 240 598 ACAATCAGGCTTCAGCTGTGGGATACCGCGGGTCAGGAACGTTTCCTAGCTCATTTCCC 657 241 AGGTACATCCGTGATTCGTCAGCTGTGATGATTTACGATATCAAAATGTTAACTCA 300 658 AGTTACATCCGTGATTCGTCAGCTGTGATGATTTACGATATCAAAATGTTAACTCA 717 301 TTCAGCAAACTCAAAAGTCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATC 360 718 TTCAGCAAACTCAAAAGTCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATC 777 361 ATCAACGCTAGTAGGAATAAGACAGATCTTGTGACAAAGGCAAGTGTCAAGTTGAGGAG 420 778 ATCAACGCTAGTAGGAATAAGACAGATCTTGTGACAAAGGCAAGTGTCAAGTTGAGGAG 837 421 GGAGAGAGAAAGCCAAAGGCTGAATTTACGTTTATTTGAAACTAGGCAAAATCGGA 480 838 GGAGAGAGAAAGCCAAAGGCTGAATTTATGTTTATTTGAAACTAGTGTGCAAAAGCTGA 897 481 TACAATGTTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTGCGGGAATGGAAGCACA 540 898 TACAATGTTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTGCGGGAATGGAAGCACA 957 541 CAGGAGGAGAGCAGAGAGACATGAGTGACATAAACTGGAAGGCAAGCTCAGAGGCAACA 600 958 CAGGAGAGAGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1017 601 GTCAGGAGAGGGGTTGTTCCCTGCTACTCTCCCATGTCATCTTCAACCCCT----- 650 1018 GTCAGTGAAGAGGGCTGTTCCCTGCTACTCTCCCATGTCATCTTCAACCCCTTCTTCAAGAGC 1077 651 --TCCTCAGAGAGCCCTTACTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 708 1078 TCCTGCTGTTGGGCCCCCTTACTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1137 709 TTCCTTTCATTAATAAGCTTTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 762			



```
Db 1138 TTCCTTCAGTAATACGTAATGCAATTCATCATCTGCTCGTGGAGA 1191

RESULT 12
LOCUS BC068486
DEFINITION Homo sapiens RAB6A, member RAS oncogene family, mRNA (cDNA clone
ACCESSION BC068486
VERSION BC068486.1 GI:46249770
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3100)
Straussberg, R.L., Collings, F.S., Wagner, L., Sherman, C.M., Schuler, G.D.,
Klausner, R.D., Zeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mulláhy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Faney, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smal, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2 (bases 1 to 3100)
Straussberg, R.
Direct Submission
Submitted (02-APR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael Brownstein / Ted Usdin
Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 168 Row: 1 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 19923230.
Location/Qualifiers
1. .3100
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:87150 IMAGE:30342562"
/tissue_type="Placenta, normal"
/clone_lib="NIH MGC_147"
/lab_host="DH10B"
/note="Vector: pBluescript"

FEATURES
source
1. .3100
/gene="RAB6A"
/note="synonym: RAB6"
/db_xref="LOCUSID:5870"
/db_xref="MIM:179513"
463. .1089
/gene="RAB6A"
/codon_start=1
/product="RAB6A protein"
/protein_id="AAH68486.1"
/db_xref="GI:46249771"
/db_xref="LOCUSID:5870"
/db_xref="MIM:179513"
/translation="MSTGGDPGNLRKFLVFLGEQSVGKTSLTRFYDSDFTYQA
TIGIDFLSKTMYLEDRITRLQTAGQERFRLSPVIRDSAAAVVVDITNNVSO
QTTWKIDVPTERGSDVILVLGNKLTDLADKROVISEGERKAKELNVMFIETSAK
YNNVQLRRVVAALPGMESTQDRSDREMDIDIKLEKPEQPVSEGGCSC"
502. .987
/misc_feature
/gene="RAB6A"
/note="RAB; Region: Rab subfamily of small GTPases"
/db_xref="CDD:cd00154"

ORIGIN
Query Match 89.8%; Score 684; DB 9; Length 3100;
Best Local Similarity 93.9%; Pred. NO. 4.4e-198;
Matches 727; Conservative 0; Mismatches 35; Indels 12; Gaps 1;
QY 1 ATGTCCGCGGCGGAGACTTCGGGAATCCGCTGAGGAATTCNAGCTGGTCTCTCGGG 60
Db 463 ATGTCCAGCGGCGGAGACTTCGGGAATCCGCTGAGGAATTCNAGCTGGTCTCTCGGG 522
QY 61 GAGCAAAACGCTTGCAAGACATCTTTGATCACCAGATTCAGTATGACAGATTTTGCACAA 120
Db 523 GAGCAAAACGCTTGCAAGACATCTTTGATCACCAGATTCAGTATGACAGATTTTGCACAA 582
QY 121 ACCTATCAGGCAATAATTTGGCATTGACATTTTATCAAAAATCATGTATCTGGAGATGA 180
Db 583 ACCTATCAGGCAATAATTTGGCATTGACATTTTATCAAAAATCATGTATCTGGAGATGA 642
QY 181 ACAATCGGCTTCGGCTGTGGGATACGCGGCTGAGGAACGCTCCGCTAGCTCATTTCCC 240
Db 643 ACAATCGGCTTCGGCTGTGGGATACGCGGCTGAGGAACGCTTCGCTAGCTCATTTCCC 702
QY 241 AGGTACATCCGTGATTTCTGCTGAGCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 703 AGTTACATCCGTGATTTCTGCTGAGCTGATGATGATGATGATGATGATGATGATGATGAT 762
QY 301 TTCCAGCAAACTACAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 763 TTCCAGCAAACTACAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822
QY 361 ATCAGCTAGTAGGAATAGAACAGATCTTCTGTCAGAGGCGAGCTCAGTTGAGGAG 420
Db 823 ATCATGCTAGTAGGAATAGAACAGATCTTCTGTCAGAGGCGAGCTCAGTTGAGGAG 882
QY 421 CGAGAGAGAAAGCCAAAGGCTGAATGTTTATGAACTTAGGCAAAAACCTGGA 480
Db 883 GGAGAGAGAAAGCCAAAGGCTGAATGTTTATGAACTTAGGCAAAAACCTGGA 942
QY 481 TACAATGTAAGAGAGCTTTTTCGAGCTGTAGCAGAGCTTTTCCGGGAATGGAAGACACA 540
Db 943 TACAATGTAAGAGAGCTTTTTCGAGCTGTAGCAGAGCTTTTCCGGGAATGGAAGACACA 1002
QY 541 CAGGACGGAACGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 1003 CAGGACGGAACGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1062
QY 601 GTCCAGAGAGGGGTTGTTCTGCTGCTCTCTCCCATGTCTTCAACCTT----- 650
Db 1063 GTCCAGAGAGGGGTTGTTCTGCTGCTCTCTCCCATGTCTTCAACCTTCTTCAAGAGC 1122
QY 651 --TCTCAGAGAGGGGTTTCTTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 708
```

Db 1123 TCACGTCTTTGGCCCCCTTACTCTTTTCATTGACTGAGTGGAATATTGGCTTGGAACCTT 1182

Qy 709 TTCCTTCATTATTAACGTTTGGCAATTCATCATCTGCTGCGCTGCTCGTGAGAGA 762

Db 1183 TTCCTTCAGTAATAACGATTATGCAATTCATCATCTGCTGCGCTGCTCGTGAGAGA 1236

# RESULT 13

AK057157

## LOCUS

DEFINITION

Homo sapiens CDNA FLJ32595 fis, clone SPLEN2000505, highly similar

to RAS-RELATED PROTEIN RAB-6.

AK057157.1 GI:16552747

VERSION oligo capping; fis (full insert sequence).

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,

Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,

Sekine, M., Ohtsuka, M., Nishi, T., Shibahara, T., Tanaka, T.,

Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,

Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,

Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,

Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,

Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,

Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,

Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,

Yosida, M., Hotsuta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,

Hara, H., Tanase, T., Nomura, Y., Togiani, S., Komai, F., Hara, R.,

Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,

Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,

Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,

Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,

Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,

Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,

Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,

Ozaki, K., Hiraoka, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,

Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K.,

Nakajima, T., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,

Oyama, H., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,

Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,

Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, K.,

Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.

Complete sequencing and characterization of 21,243 full-length

human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)

14702039

2

Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R.,

Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J.,

Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K.,

Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H.,

Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,

Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahara, K.,

Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

Unpublished

3 (bases 1 to 3120)

Isogai, T., Otsuki, T. and Sugiyama, T.

Direct Submission

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana Kisarazu, Chiba 292-0812, Japan

(E-mail: gcnomic@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI); (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

## FEATURES

source

Location/Qualifiers

1..3120

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="SPLEN2000505"

/tissue type="spleen"

/clone\_lib="SPLEN2"

/note="Cloning vector: pME18SFL3"

487..1113

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB71371.1"

/db\_xref="GI:16552748"

/translation="MSTGGDFGNPLRKFVLFLGEQSVGKTSILITREWYDSFDNTYQA

TIGDIFLSKTYLEDRTIRLQLDWDTAGQFRSLIPSYIRDSAAAIVVYDITWNVSFO

QTKWIDDVTERGSDVILMVGKNTDLADKRQVSIIEGGRKAKELNMFIEISAKAG

YINVKQLFRVVAALPGMESTQDRSREDMIDIKLEKPEQPVSEGGCSC"

## ORIGIN

Query Match 89.8%; Score 684; DB 9; Length 3120;

Best Local Similarity 93.9%; Pred. No. 4.4e-198; Indels 12; Gaps 1;

Matches 72; Conservative 0; Mismatches 35;

Qy 1 ATGTCCCGCGGCGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60

Db 487 ATGTCCCGCGGCGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 546

Qy 61 GAGCAAAACGTTGCAAAAGACATCTTGTATCACCAGATTCAGGTATGACAGTTTGCACAAC 120

Db 547 GAGCAAAACGTTGCAAAAGACATCTTGTATCACCAGATTCAGGTATGACAGTTTGCACAAC 606

Qy 121 ACCTATCAGCAATAATTTGGCATTGACCTTTTATCAAAAATATGATCTTGGAGGATGGA 180

Db 607 ACCTATCAGCAATAATTTGGCATTGACCTTTTATCAAAAATATGATCTTGGAGGATGGA 666

Qy 181 ACATCGGGCTTCGGCTGTGGGATACCGCGGGTCAGGAAAGCTCTCCGTAGCCTCATTTCCC 240

Db 667 ACATCGGGCTTCGGCTGTGGGATACCGCGGGTCAGGAAAGCTCTCCGTAGCCTCATTTCCC 726

Qy 241 AGGTACATCGTGTATCTGCTGACGCTGTAGTAGTTTACGATATCAGAAATGTTAACTCA 300

Db 727 AGTTTACATCGTGTATCTGCTGACGCTGTAGTAGTTTACGATATCAGAAATGTTAACTCA 786

Qy 301 TTCAGCAAACTACAAAGTCGATTGATGTGACAGACAGAAAGAGGAGTGTATTATC 360

Db 787 TTCAGCAAACTACAAAGTCGATTGATGTGACAGACAGAAAGAGGAGTGTATTATC 846

Qy 361 ATCAGCGTAGTAGGAAATAGAACAGATCTTGTGACAGAGGCAAGTGTCAAGTTGAGGAG 420

Db 847 ATCAGCGTAGTAGGAAATAGAACAGATCTTGTGACAGAGGCAAGTGTCAAGTTGAGGAG 906

Qy 421 GGAGAGAGGAAAGCAAAAGCGCTGAATGTTCGTTTATGAAATAGGCGCAAAAGCTGGA 480

Db 907 GGAGAGAGGAAAGCAAAAGCGCTGAATGTTCGTTTATGAAATAGGCGCAAAAGCTGGA 966

Qy 481 TACAATCTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTCCCGGGAATGGAAGACACA 540

Db 967 TACAATCTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTCCCGGGAATGGAAGACACA 1026

Qy 541 CAGGACGGAAGCAGAGAGACATGAGTGACATATAAACTGGAAGACCTTCAGGAGCAAAACA 600

Db 1027 CAGGACGGAAGCAGAGAGACATGAGTGACATATAAACTGGAAGACCTTCAGGAGCAAAACA 1086

Qy 601 GTCCAGCAAGGGGTGTCTCTGCTACTCTCCCATGTCTCAACCTCTTCAACCTCTTCAACCTCT 650

Db 1087 GTCCAGCAAGGGGTGTCTCTGCTACTCTCCCATGTCTCAACCTCTTCAACCTCTTCAACCTCT 1146

Qy 651 -TCCTCAGAAAGCCCTTACTCTTTCATTGATGCTGAGTGTGAATATTGGCTTGAACCTT 708

Db	1147	TCAGTCTTTGGCCCCCTTACTCTTTTCATTGATCGAGTGTGAATATTCGCTTGAACCTT	1206
Qy	709	TTCCCTTCAATTAATACGTTTTCGCAATTCATTCATTCCTGCCCTGCTCGTGGAGA	762
Db	1207	TTCCCTTCAGTAATAACGATTGCAATTCATCATTCCTGCTGCCCTGCTCGTGGAGA	1260
RESULT 14			
LOCUS	AR447780	3745 bp	DNA linear PAT 20-FEB-2004
DEFINITION	Sequence 387 from patent US 6673549.		
ACCESSION	AR447780		
VERSION	AR447780.1	GI:42676104	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 3745)		
AUTHORS	Furness,L.M. and Buchbinder,J.L.		
TITLE	Genes expressed in C3A liver cell cultures treated with steroids		
JOURNAL	Patent: US 6673549-A 387 06-JAN-2004;		
FEATURES	Location/Qualifiers		
source	1..3745		
ORIGIN	/organism="unknown"		
	/mol_type="genomic DNA"		
Query Match	83.9%;	Score 639.2;	DB 6; Length 3745;
Best Local Similarity	90.3%;	Pred. No. 2.5e-184;	
Matches	699;	Conservative 0;	Mismatches 63; Indels 12; Gaps 1;
Qy	1	ATGTCGGCGGGGAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGGTGTCTCTGGGG	60
Db	494	ATGTCACCGGGGGAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGGTGTCTCTGGGG	553
Qy	61	GAGCAAGCGTTGCCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCGACAC	120
Db	554	GAGCAAGCGTTGGAAAGACATCTTTGATCACCAGATTCATGTATGACAGTTTTCGACAC	613
Qy	121	ACCTATCAGCAATTAATTCGCAATTCGCTTTTATCACCAGATTCAGGTATGACAGTTTTCGACAC	180
Db	614	ACCTATCAGCAATTAATTCGCAATTCGCTTTTATCACCAGATTCAGGTATGACAGTTTTCGACAC	673
Qy	181	ACAATCGGCTTCGGCTGGGATACGGGGTTCAGGAACGCTCCGTAGCGCTCATTCCTC	240
Db	674	ACAGTACGATTGCAATATTGGACACAGCAGGTCAAGCGGTTGAGGAGCTTGATTCCT	733
Qy	241	AGGTATACCTCGTATTCGCTGCAGCTGTAGTGTATTCAGTATCACAAATGTTAACTCA	300
Db	734	AGCTACATTCGTCGACTCCACTGGCAGTGTGTTTATGATATCACAAATGTTAACTCA	793
Qy	301	TTCCAGCAAACTACAAAGTGGATGTATGATGTACAGAACAGAAAGGAAAGTGTATTC	360
Db	794	TTCCAGCAAACTACAAAGTGGATGTATGATGTACAGAACAGAAAGGAAAGTGTATTC	853
Qy	361	ATCAGCTAGTAGGAATAGACAGATCTTGTCTGACAGAGCAAGTGTACGTTGAGGAG	420
Db	854	ATCATGCTAGTAGGAATATAAACAAGATCTTGTCTGACAGAGCAAGTGTCAATGAGGAG	913
Qy	421	GGAGAGAGAAAGCCAAAGGGCTGAATGTAGTGTATTGAACTAGGCAAAAACCTGGA	480
Db	914	GGAGAGAGAAAGCCAAAGAGCTGAATGTATGTTTATTGAACTAGTGTCAAAAGCTGA	973
Qy	481	TACAATGTAAAGCAGCTTTTCGAGTGTAGCAGCAGCTTTCCCGGGAATGAAAGCACA	540
Db	974	TACAATGTAAAGCAGCTTTTCGAGTGTAGCAGCAGCTTTCCCGGGAATGAAAGCACA	1033
Qy	541	CAGGACGGAGCAGAGACATGATGATCAATAAACTGGAAGCCCTCAGGAGCAACA	600
Db	1034	CAGGACAGAGCAGAGAGATGATGATCAATAAACTGGAAGCCCTCAGGAGCAACA	1093
Qy	601	GTCAAGCAAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCTCT	650
Db	1094	GTCAAGCAAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCTCTTCAAGC	1153
Qy	651	--TCTCAGAGAGCCCCCTTACTCTTTTCATTGATCGAGTGTGAATATTCGCTTGAACCTT	708
Db	1154	TCATGCTTTGGCCCCCTTACTCTTTTCATTGATCGAGTGTGAATATTCGCTTGAACCTT	1213
Qy	709	TTCCCTTCATTAATAACGTTTTCGCAATTCATTCATTCCTGCCCTGCTCGTGGAGA	762
Db	1214	TTCCCTTCAGTAATAACGATTGCAATTCATTCATTCCTGCCCTGCTCGTGGAGA	1267
RESULT 15			
LOCUS	BC019118	1770 bp	mRNA linear ROD 04-OCT-2003
DEFINITION	Mus musculus RAB6, member RAS oncogene family, mRNA (cdna clone		
ACCESSION	MGC:29434 IMAGE:3711087), complete cds.		
VERSION	BC019118		
KEYWORDS	BC019118.1	GI:17512289	
SOURCE	MGC.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	1 (bases 1 to 1770)		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	Klausner,R.D., Collins,P.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalley,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 1770)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Jeffrey Green M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a> Contact: <a href="mailto:nisc_mgc@nhgri.nih.gov">nisc_mgc@nhgri.nih.gov</a> Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgenev,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,		

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 39 Row: a Column: 19  
This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.

## FEATURES

Location/Qualifiers

1..1770

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="MGC:29434 IMAGE:3711087"

/tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."

/clone\_lib="NCI CGAP\_Mam6"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

## gene

1..1770

/gene="Rab6"

/db\_xref="LocusID:19346"

/db\_xref="MGI:894313"

271..897

/gene="Rab6"

/codon\_start=1

/product="Rab6 protein"

/protein\_id="AAH19118.1"

/db\_xref="GI:17512290"

/db\_xref="LocusID:19346"

/db\_xref="MGI:894313"

/translation="MSGAGDRGNELRFLVFLGEQSVGKSLITRMVDSFDNTYQA  
TIGIDFLSKMYLEDRTIQLWDTAGDERPSLIPSYIRDSAAVYVDITNVNSFQ  
QTWIDVDVTRGSDVILVILADKRVQSIIEGKAKELNMFIEISAKAG  
YNVQLFRVAAALPGMESTQDRSDMDIKLEKPOEQPVNEGSCS"

310..792

/gene="Rab6"

/note="RAB; Region: Rab subfamily of small GTPases"

/db\_xref="CDD:smart00175"

## ORIGIN

Query Match 83.8%; Score 638.2; DB 10; Length 1770;

Best Local Similarity 90.3%; Pred. No. 4.6e-184; Mismatches 63; Indels 12; Gaps 1;

Matches 698; Conservative 0;

Qy 1 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTCTTCCTGGGG 60

Db 271 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTCTTCCTGGGA 330

Qy 61 GAGCAAGCGTTGCAAGACATCTTGTATCACCAGATTCAGGTATGACAGTTTTCACAAAC 120

Db 331 GAGCAGAGCGTTGGAAGAGACGCTCTTGATCATCCCGAATTCATGTATGACAGTTTTCACAAAC 390

Qy 121 ACCTATCAGGCAATTAATTTGGCATTCACCTTTTATCAAAAACATGTACTTCGAGGATGGA 180

Db 391 ACCTATCAGGCAACAAATTTGGCATTCACCTTTTATCAAAAACAAATGTACTTCGAGGATGGA 450

Qy 181 ACAATTCGGGCTTCGGCTCTGGGATACGGGGGTTCAGGAACGCTCTCCGTAGCCTCATTTCCC 240

Db 451 ACCATCAGGCTGCAGCTCTGGGATACGGGGGTTCAGGAACGCTCTCCGTAGCCTCATTTCCC 510

Qy 241 AGGTATACCTCGTGATCTTGCTGCAGCTGTAGTTTACGATATACAAATGTTAACTCA 300

Db 511 AGTTATACCTCGTGATCTTGCTGCAGCTGTAGTTTACGATATACAAATGTTAACTCA 570

Qy 301 TTCCAGCAAACTACAAAGTGGATTCATGTCTCAGAACAGAGAGGAAGTGTGTTATC 360

Db 571 TTCAGCAAACTACAAATGGAATTCATGTCTCAGAACAGAGAGGAAGTGTGTTATC 630

Qy 361 ATCAGCGTAGTAGGAAATAGAACAGATCTTGCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420

Db 631 ATCATGCTAGTAGGAAATAGAACAGATCTTGCTGATTAAGAGGCAAGTGTCTCAATTTGAGGAG 690

Qy 421 GGAGAGAGGAAGCCAAAGGGCTGAATGTTTACGTTTATTGAAACTAGGCAGAAAACCTGGA 480  
Db 691 GGAGAGAGGAAGCCAAAGAGCTGAATGTTTATTGAAACCACTGCAAAAGCAGGA 750  
Qy 481 TACAATGTAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTTGGCGGGAATGGAAAGCACA 540  
Db 751 TACAATGTAAGCAGCTCTTTTCGCGCGTGTGCTGCAGCTTTACCTGGAATGGAAAGCACA 810  
Qy 541 CAGGACGGAAGCAGAGAGACATGATGACATAAAACTGGAAGACCTTCAGGAGCAACA 600  
Db 811 CAGGACGGAAGCAGAGAGACATGATGACATAAAACTGGAAGACCTTCAGGAGCAACA 870  
Qy 601 GTCAATGTAAGCAGCTCTTTTCCTGCTACTCTCCCATGTCATCTTCAACCTTCTCTCAGAAG 660  
Db 871 GTCAATGTAAGCAGCTCTTTTCCTGCTACTCTCCCATGTCATCTTCAACCTTCTCTCAGAAG 930  
Qy 661 C-----CCCCCTTACTCTTTCAATTGACTGTCAGTGTGAATATTGGCTTGAACCTT 708  
Db 931 CTCACCTGCTTTGGCCCCCATACTCTTTTCATTGACTGTCAGTGTGAATATTGGCTTGAACCTT 990  
Qy 709 TTCCCTTCATTAATAACGTTTTCGCAATTTCATCTGCTGCTGCTCTCTCTGCTGAG 761  
Db 991 TTCCCTTCAGTAATAACGTTTTCGCAATTTCATCTGCTGCTGCTGCTGCTGCTGCTGAG 1043

Search completed: April 25, 2005, 10:41:08

Job time : 3402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 05:41:05 ; Search time 475 Seconds  
(without alignments)  
9496.511 Million cell updates/sec

Title: US-09-441-857-11

Perfect score: 762

Sequence: 1 atgtccggcggcgagactt.....tgcgtcgtctcgtggaga 762

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	762	100.0	762	3	Aaz93839
2	761	99.9	3084	5	Abx71303
C 3	759.4	99.7	349981	10	Adc87619 Human GPC
C 4	759.4	99.7	349989	10	Adc86916 Human GPC
5	684	89.8	1460	3	Aaz93840
6	684	89.8	3079	12	Adn03693
7	672	88.2	719	3	Aaz93834
8	639.2	83.9	3745	12	Adl12658
9	637.8	83.7	3118	3	Aaf15989
10	609.2	79.9	739	3	Aaz93836
11	585.6	76.9	624	3	Aaz93841
12	559.6	73.4	718	3	Aaz93835
13	559.6	73.4	740	3	Aaz93837
14	526	69.0	573	3	Aaz93838
15	489.8	64.3	3195	5	Aas87692
16	401	52.6	2456	10	Ada52862
17	384.8	50.5	1219	4	Aak52711
18	384.8	50.5	1461	4	Aak51727
19	383.2	50.3	1266	12	Adp07320
20	383.2	50.3	1266	13	Adr25795

21	383.2	50.3	1589	2	Aaz11736
22	382	50.1	443	8	Abx42190
23	352.2	46.2	3826	10	Adc86930
24	341.8	44.9	424	9	ACH18328
25	296.4	38.9	1188	4	ABL11393
C 26	296.4	38.9	3188	4	ABL11392
C 27	286.8	37.6	477	5	ABA20123
28	276.8	36.3	642	5	AAS87691
29	268.4	35.2	1074	3	AAC33337
30	251.6	33.0	523	13	ACN59588
31	250	32.8	636	10	ADC76216
32	250	32.8	636	10	ADK55884
33	247	32.4	1355	3	AAC46575
34	246.2	32.3	624	3	AAC42719
35	245.2	32.2	928	6	ABI99630
36	244.4	32.1	926	3	AAP14485
37	236	31.0	495	3	AAC41466
38	227	29.8	646	10	ADD16517
39	227	29.8	646	11	ADM44783
40	227	29.8	647	10	ADC75668
41	227	29.8	723	10	ADC76194
42	227	29.8	723	10	ADK58490
43	216	28.3	574	13	ADR65698
44	213	28.0	645	3	AAC43018
45	211.8	27.8	791	13	ADR65699

ALIGNMENTS

RESULT 1

Aaz93839

ID Aaz93839 standard; DNA; 762 BP.

XX

AC Aaz93839;

XX 29-AUG-2000 (first entry)

XX WTH3 coding sequence.

XX WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;

XX antibody; immunogen; mutation; detection; therapy; human; ss.

OS Homo sapiens.

XX Key

XX Location/Qualifiers

FT CDS

FT 1..762

FT /\*tag= a

FT /product= "WTH3"

XX WO200029625-A1.

XX PD

XX 25-MAY-2000.

XX 18-NOV-1999; 99WO-US027630.

XX PR 18-NOV-1998; 98US-0108994P.

XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;

XX WPI; 2000-387828/33.

XX P-PSDB; AAY83649.

XX Novel nucleic acids encoding wth3 protein useful for inhibiting multiple

XX drug resistance in tumor cells for identifying the modulators of drug

XX resistance.

XX Claim 3; Page 62-64; 82pp; English.

XX Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)

XX Differential Display (MDD) methods and can be used in methods for

XX CC

CC detecting methylation patterns in multiple drug resistance loci. Genes  
CC are frequently not methylated in cells where they are expressed but are  
CC methylated in cell types where they are not expressed. Tumour cell DNA is  
CC often methylated to a different extent and in different regions when  
CC compared to DNA of normal cells. The methylation pattern in a multiple  
CC drug resistance locus can be altered and give rise to altered expression  
CC patterns of that multiple drug resistance locus. Nucleic acids  
CC corresponding to the identified loci such as WH3 and RAB6 nucleic acids  
CC can be used as probes for detecting mutations and methylation patterns of  
CC those loci. The nucleic acids and their homologues are useful for  
CC inhibition of multiple drug resistance and for treating tumors exhibiting  
CC multiple drug resistance. They are also useful for detecting and  
CC measuring the expression of mRNA from identified genes and for  
CC determining suitable therapeutic treatment. Antibodies directed against  
CC immunogenic fragments of WH3 and RAB6 are useful for detecting specific  
CC proteins and polypeptides in tissues or body fluids of patients  
XX  
SQ Sequence 762 BP; 221 A; 157 C; 194 G; 190 T; 0 U; 0 Other;

Query Match 100.0%; Score 762; DB 3; Length 762;  
Best Local Similarity 100.0%; Pred. No. 4.9e-227;  
Matches 762; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60  
DB |||||||  
QY 1 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60  
DB |||||||

QY 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGGACAAC 120  
DB |||||||

QY 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGGACAAC 120  
DB |||||||

QY 121 ACCTATCAGGCAATTAATTTGGCATTTGACATTTTATCAAAAATCTGCTGAGGATGGA 180  
DB |||||||

QY 121 ACCTATCAGGCAATTAATTTGGCATTTGACATTTTATCAAAAATCTGCTGAGGATGGA 180  
DB |||||||

QY 181 ACAATCGGGCTTCGGCTGTGGGATACGGGGTCAGGAAACGTCCTCGTAGCTCATTTCCC 240  
DB |||||||

QY 181 ACAATCGGGCTTCGGCTGTGGGATACGGGGTCAGGAAACGTCCTCGTAGCTCATTTCCC 240  
DB |||||||

QY 241 AGGTACATCCGTTGATTTCTGTCAGCTGTAGTTTACGATATACAAAATGTTAACTCA 300  
DB |||||||

QY 241 AGGTACATCCGTTGATTTCTGTCAGCTGTAGTTTACGATATACAAAATGTTAACTCA 300  
DB |||||||

QY 301 TTCAGCAAACTACAAAGTGGATTCATGATCTCAGAACAGAAAGGAGTGTGTTATC 360  
DB |||||||

QY 301 TTCAGCAAACTACAAAGTGGATTCATGATCTCAGAACAGAAAGGAGTGTGTTATC 360  
DB |||||||

QY 361 ATCAGCTAGTAGGAATAGAAACAGATCTTGTGACAAAGGCAAGTGTCTAGTTGAGGAG 420  
DB |||||||

QY 361 ATCAGCTAGTAGGAATAGAAACAGATCTTGTGACAAAGGCAAGTGTCTAGTTGAGGAG 420  
DB |||||||

QY 421 GGAGAGAGAAAGCCAAAGGGCTGAATGTTATGAAATTCAGGCAAAATCTGGA 480  
DB |||||||

QY 421 GGAGAGAGAAAGCCAAAGGGCTGAATGTTATGAAATTCAGGCAAAATCTGGA 480  
DB |||||||

QY 481 TACATGTAAAGCAGCTCTTTGAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGCACA 540  
DB |||||||

QY 481 TACATGTAAAGCAGCTCTTTGAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGCACA 540  
DB |||||||

QY 541 CAGGACGGAAGCAGAGAGACATGATGACATAAAATCTGGAAGAGCCCTCAGGAGCAACA 600  
DB |||||||

QY 541 CAGGACGGAAGCAGAGAGACATGATGACATAAAATCTGGAAGAGCCCTCAGGAGCAACA 600  
DB |||||||

QY 601 GTCAGCAAGGGGGTGTTCCTGCTACTCTCCCATGTCACTTCAACCCCTTCCTCAGAAG 660  
DB |||||||

QY 601 GTCAGCAAGGGGGTGTTCCTGCTACTCTCCCATGTCACTTCAACCCCTTCCTCAGAAG 660  
DB |||||||

QY 661 CCCCTTACTCTTTCATTTGACTGCTGATGTAATTTGGCTTTGAACTTTTCCCTTCATTA 720  
DB |||||||

QY 661 CCCCTTACTCTTTCATTTGACTGCTGATGTAATTTGGCTTTGAACTTTTCCCTTCATTA 720  
DB |||||||

QY 721 ATAAGCTTTTGCATTCATCATTTGCTGCTGTCTCGTGGAGA 762  
DB |||||||

Db 721 ATAAGCTTTTGCATTCATCATTCATTCGCTGTCTCGTGGAGA 762

RESULT 2  
ABX71303  
ID ABX71303 standard; cDNA; 3084 BP.  
XX AC ABX71303;  
XX AC ABX71303;  
XX 14-APR-2003 (first entry)  
XX Human intracellular transport and trafficking cDNA DKFZphfk2\_4k14.  
DE Human; gene; gene therapy; vaccine; disease treatment; detection; ss.  
KW Homo sapiens.  
XX OS Homo sapiens.  
XX PN WO200112659-A2.  
XX 22-FEB-2001.  
XX 18-AUG-2000; 2000WO-IB001496.  
XX 18-AUG-1999; 99US-0149499P.  
XX 28-SEP-1999; 99US-0156503P.  
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX WIemann S;  
XX WPI; 2001-327840/34.  
XX P-PSDB; ABUS2844.  
XX Nucleic acids having the sequences of clones isolated from libraries of  
XX different human tissues, useful in recombinant DNA methodologies.  
XX Claim 1; Page 432-433; 1095pp; English.  
XX This invention describes novel polynucleotides and polypeptides isolated  
XX from human cDNA libraries which can be used for gene therapy or in  
XX vaccines. The polynucleotides of the invention and antibodies encoded by  
XX them may be used in the prevention, diagnosis and treatment of diseases  
XX associated with inappropriate polypeptide expression. The products of the  
XX invention may also be used to identify modulators of expression and  
XX activity and to down regulate expression and activity. The antibodies of  
XX the invention may also be used as diagnostic agents for detecting the  
XX presence of polypeptides in samples. This sequence encodes a polypeptide  
XX described in the disclosure of the invention  
XX SQ Sequence 3084 BP; 881 A; 565 C; 662 G; 876 T; 0 U; 0 Other;

Query Match 99.9%; Score 761; DB 5; Length 3084;  
Best Local Similarity 100.0%; Pred. No. 2.1e-226;  
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60  
DB |||||||  
QY 456 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 515  
DB |||||||

QY 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGGACAAC 120  
DB |||||||

QY 516 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGGACAAC 575  
DB |||||||

QY 121 ACCTATCAGGCAATTAATTTGGCATTTGACATTTTATCAAAAATCTGCTGAGGATGGA 180  
DB |||||||

QY 576 ACCTATCAGGCAATTAATTTGGCATTTGACATTTTATCAAAAATCTGCTGAGGATGGA 635  
DB |||||||

QY 181 ACATCGGGCTTCGGCTGTGGGATACGGGGTCAGGAAACGTCCTCGTAGCTCATTTCCC 240  
DB |||||||

QY 636 ACATCGGGCTTCGGCTGTGGGATACGGGGTCAGGAAACGTCCTCGTAGCTCATTTCCC 695  
DB |||||||

QY 241 AGGTACATCCGTTGATTCATTCATTTGCTGCTGTAGTATGATATCAAAATGTTAACTCA 300  
DB |||||||

Dd 696 AGGTACATCCGTGATTCTCTGCTGAGCTGTAGTATTACGATATACAAATGTTAACTCA 755  
Qy 301 TTCCAGCAAACTACAAAGTGGATTGATGATGTGAGAAACAGAAAGAGGAGTGTATTC 360  
Dd 756 TTCCAGCAAACTACAAAGTGGATTGATGATGTGAGAAACAGAAAGAGGAGTGTATTC 815  
Qy 361 ATCAGCTAGTAGGAAATAGAACAGATCTTGCTGACAGAGGCAAGTGTCACTGAGGAG 420  
Dd 816 ATCAGCTAGTAGGAAATAGAACAGATCTTGCTGACAGAGGCAAGTGTCACTGAGGAG 875  
Qy 421 CGAGAGAGAAAGCCAAAGGGCTGAATGTTAGCTTTATTGAACTAGGGCAAAACTGGA 480  
Dd 876 GGAGAGAGAAAGCCAAAGGGCTGAATGTTAGCTTTATTGAACTAGGGCAAAACTGGA 935  
Qy 481 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTTCCGGGAAATGGAAGCACA 540  
Dd 936 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTTCCGGGAAATGGAAGCACA 995  
Qy 541 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Dd 996 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1055  
Qy 601 GTGAGGAGAGGGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Dd 1056 GTGAGGAGAGGGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115  
Qy 661 CCCCTTACTCTTTCATGACTGCAAGTGTGAATATTGGCTTGAACCTTTTCCTCTTCAATTA 720  
Dd 1116 CCCCTTACTCTTTCATGACTGCAAGTGTGAATATTGGCTTGAACCTTTTCCTCTTCAATTA 1175  
Qy 721 ATAAGCTTTTGAATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761  
Dd 1176 ATAAGCTTTTGAATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1216

## RESULT 3

ADC87619/c  
ID ADC87619 standard; DNA; 349981 BP.  
XX  
AC ADC87619;  
XX  
DT 01-JAN-2004 (first entry)  
DE Human GPCR related polynucleotide SEQ ID NO:2072.  
XX  
KW ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;  
KW gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1270724-A2.  
XX  
PD 02-JAN-2003.  
XX  
PF 18-JUN-2002; 2002BP-00013517.  
XX  
PR 18-JUN-2001; 2001JP-00246789.  
XX  
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
XX WPI; 2003-315783/31.  
XX  
PT New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
PS Disclosure; SEQ ID NO 2072; 28pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of

CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The protein  
CC sequences shown in ADC87618-ADC87623 represent polynucleotide sequences  
CC related to the invention.

XX SQ Sequence 349981 BP; 100900 A; 78560 C; 77396 G; 92823 T; 0 U; 302 Other;  
Query Match 99.7%; Score 759.4; DB 10; Length 349981;  
Best Local Similarity 99.9%; Pred No. 7.3e-225;  
Matches 760; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTCGTGTCTCTGGGG 60  
Dd 45699 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTCGTGTCTCTGGGG 45640  
Qy 61 GAGCAACCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCACAC 120  
Dd 45639 GAGCAACCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCACAC 45580  
Qy 121 ACCTATCAGGCAATAATTGGCATTGACCTTTTATCAAAATCTATGCTTGGAGATGGA 180  
Dd 45579 ACCTATCAGGCAATAATTGGCATTGACCTTTTATCAAAATCTATGCTTGGAGATGGA 45520  
Qy 181 ACAATCGGGCTTCGGCTGTGGGATACGGCGGTGAGAACGCTCTCCGTAGCTCTATCCC 240  
Dd 45519 ACAATCGGGCTTCGGCTGTGGGATACGGCGGTGAGAACGCTCTCCGTAGCTCTATCCC 45460  
Qy 241 AGGTACATCCGTGATTCTGCTGAGCTGTAGTATTAGGATATACAAATGTTAACTCA 300  
Dd 45459 AGGTACATCCGTGATTCTGCTGAGCTGTAGTATTAGGATATACAAATGTTAACTCA 45400  
Qy 301 TTCCAGCAAACTACAAAGTGGATTGATGATGTGAGAAAGAGGAGTGTATTC 360  
Dd 45399 TTCCAGCAAACTACAAAGTGGATTGATGATGTGAGAAAGAGGAGTGTATTC 45340  
Qy 361 ATCAGCTAGTAGGAAATAGAACAGATCTTCTGCAAGAGCAAGTGTCACTGAGGAG 420  
Dd 45339 ATCAGCTAGTAGGAAATAGAACAGATCTTCTGCAAGAGCAAGTGTCACTGAGGAG 45280  
Qy 421 CGAGAGAGAAAGCCAAAGGGCTGAACTGTTAGCTTTATTGAACTAGGGCAAAAGCTGA 480  
Dd 45279 CGAGAGAGAAAGCCAAAGGGCTGAACTGTTAGCTTTATTGAACTAGGGCAAAAGCTGA 45220  
Qy 481 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTTCCGGGAAATGGAAGCACA 540  
Dd 45219 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTTCCGGGAAATGGAAGCACA 45160  
Qy 541 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Dd 45159 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 45100  
Qy 601 GTCAGGAGAGGGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Dd 45099 GTCAGGAGAGGGGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 45040  
Qy 661 CCCCTTACTCTTTCATGACTGCAAGTGTGAATATTGGCTTGAACCTTTTCCTCTTCAATTA 720  
Dd 45039 CCCCTTACTCTTTCATGACTGCAAGTGTGAATATTGGCTTGAACCTTTTCCTCTTCAATTA 44980  
Qy 721 ATAAGCTTTTGAATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761  
Dd 44979 ATAAGCTTTTGAATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 44939

RESULT 4  
ADC86916/c  
ID ADC86916 standard; DNA; 349989 BP.  
XX  
AC ADC86916;  
XX  
XX 01-JAN-2004 (first entry)  
DT



XX Human GPCR gene SEQ ID NO:1369.  
DE ds; gene; human; GPCR;  
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.  
XX Homo sapiens.  
OS  
PN EP1270724-A2.  
XX  
XX 02-JAN-2003.  
XX  
XX 18-JUN-2002; 2002EP-00013517.  
XX  
XX 18-JUN-2001; 2001JP-00246789.  
XX  
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
XX WPI; 2003-315783/31.  
DR P-PSDB; ADC86917.  
XX  
XX New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
XX Claim 1; SEQ ID NO 1369; 28pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.  
XX  
XX Sequence 349989 BP; 93096 A; 80097 C; 83967 G; 92327 T; 0 U; 502 Other;  
SQ  
Query Match 99.7%; Score 759.4; DB 10; Length 349989;  
Best Local Similarity 99.9%; Pred. No. 7,3e-225;  
Matches 760; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTGTCTCTGGGG 60  
Db ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTGTCTCTGGGG 345649  
QY 61 GAGCAAAAGCGTTGCAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTTCAGCAAC 120  
Db GAGCAAAAGCGTTGCAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTTCAGCAAC 345589  
QY 121 ACCTATCAGGCAATAATTGGCAATTCACATTTTATCAAAAACATATGCTACCTGGAGATGGA 180  
Db ACCTATCAGGCAATAATTGGCAATTCACATTTTATCAAAAACATATGCTACCTGGAGATGGA 345529  
QY 181 ACAATCGGCTTCGGCTGTGGGATACGGGGGTGAGGAACGCTCTCCGTAGCCTCATTTCCC 240  
Db ACAATCGGCTTCGGCTGTGGGATACGGGGGTGAGGAACGCTCTCCGTAGCCTCATTTCCC 345469  
QY 241 AGGTACATCCGTGATCTCGTGCAGCTGTAGTATGATGATACAGTATACAAATGTTAACTCA 300  
Db AGGTACATCCGTGATCTCGTGCAGCTGTAGTATGATGATACAGTATACAAATGTTAACTCA 345409  
QY 301 TTCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGGAGGATGTTATC 360  
Db TTCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGGAGGATGTTATC 345349  
QY 361 ATCAGCTAGTAGGAAATAGAACAGATCTTGTGTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420  
Db ATCAGCTAGTAGGAAATAGAACAGATCTTGTGTGCAAGAGGCAAGTGTCTAGTTGAGGAG 345289

QY 421 GGAGAGGAGAAAGCAAAAGGGCTGAATGCTTTACGTTTATTGAACTAGGGCAAAACTGGA 480  
Db GGAGAGGAGAAAGCAAAAGGGCTGAATGCTTTACGTTTATTGAACTAGGGCAAAAGCTGGA 345229  
QY 481 TACAATGTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTGCGGGGAATGGAAGCACA 540  
Db TACAATGTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTGCGGGGAATGGAAGCACA 345169  
QY 541 CAGACGGAAGCAGAGAAAGACATGAGTGACATATAAACTGGAAGACCTCAGGAGCAACA 600  
Db CAGACGGAAGCAGAGAAAGACATGAGTGACATATAAACTGGAAGACCTCAGGAGCAACA 345109  
QY 601 GTCAGCGAAGGGGTGTGTTCTCTACTCTCCCATGTCATCTCAACCCCTTCCTCAGAAG 660  
Db GTCAGCGAAGGGGTGTGTTCTCTACTCTCCCATGTCATCTCAACCCCTTCCTCAGAAG 345049  
QY 661 CCCCTTACTCTTTTCATTGACTGCAGTGTAATATGGCTTTGAACCTTTCCCTTCATTA 720  
Db CCCCTTACTCTTTTCATTGACTGCAGTGTAATATGGCTTTGAACCTTTCCCTTCATTA 344989  
QY 721 ATAAAGTTTGCATTTTCATTGACTGCAGTGTAATATGGCTTTTCCTGTGGAG 761  
Db ATAAAGTTTGCATTTTCATTGACTGCAGTGTAATATGGCTTTTCCTGTGGAG 344948  
RESULT 5  
AAZ93840  
ID AAZ93840 standard; DNA; 1460 BP.  
XX  
AC AAZ93840;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE RAB6C coding sequence.  
XX  
KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
KW antibody; immunogen; mutation; detection; therapy; human; ss.  
XX  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 420..1046  
FT /\*tag= a  
FT /product= "RAB6C"  
XX  
XX WO200029625-A1.  
XX  
XX 25-MAY-2000.  
XX  
XX 18-NOV-1999; 99WO-US027630.  
XX  
XX 18-NOV-1998; 98US-0108994P.  
XX  
XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.  
XX  
XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;  
XX  
XX WPI; 2000-387828/33.  
DR P-PSDB; AAY83403.  
XX  
XX Novel nucleic acids encoding wth3 protein useful for inhibiting multiple  
PT drug resistance in tumor cells for identifying the modulators of drug  
PT resistance.  
XX  
XX Claim 6; Page 68-70; 82pp; English.  
XX  
XX Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)  
CC Differential Display (MDD) methods and can be used in methods for  
CC detecting methylation patterns in multiple drug resistance loci. Genes  
CC are frequently not methylated in cells where they are expressed but are  
CC methylated in cell types where they are not expressed. Tumour cell DNA is  
CC often methylated to a different extent and in different regions when  
CC compared to DNA of normal cells. The methylation pattern in a multiple



ADN03693

QY 361 ATCACGCTAGTAGGAATAGAACAGATCTTGCTCACAGAGGCAAGTGTCAGTTGAGGAG 420

QY  
364 ATCACCCCTAGTAGGHHHATAGAAACAGATCCTTCTGACACAGGCGAAAGTCTCAATGAGGAG  
787 ATCATGCTAGTAGGAAATAAAACAGATCTTCTGCTGACACAGGCGAAAGTCTCAATGAGGAG



XX 06-MAY-2004 (first entry)  
XX Human steroid-induced C3A liver cell cDNA #387.  
XX ss; gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;  
XX steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.  
XX  
XX Homo sapiens.  
XX US6673549-B1.  
XX  
XX 06-JAN-2004.  
XX 12-OCT-2001; 2001US-00976594.  
XX 12-OCT-2000; 2000US-0240409P.  
XX (INCY-) INCYTE CORP.  
XX Furness LM, Buchbinder JL;  
XX WPI; 2004-068610/07.  
XX  
XX Combination useful for preparing a composition for treating liver  
XX disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,  
XX comprises cDNAs that are differentially expressed in response to steroid  
XX treatment.  
XX  
XX Claim 1; SEQ ID NO 387; 141pp; English.  
XX  
XX The invention relates to a combination comprising cDNAs that are  
XX differentially expressed in response to steroid treatment. Also included  
XX are the following: a high throughput method for using a cDNA to detect  
XX differential expression of nucleic acids in a sample; and a high  
XX throughput method of screening molecules or compounds to identify a  
XX ligand that specifically binds a cDNA. The sample is from a subject with  
XX Wilson disease and comparison of a standard defines a stage of that  
XX disease. The high throughput method of screening molecules or compounds  
XX to identify a ligand that specifically binds a cDNA comprises: combining  
XX the combination with molecules or compounds under conditions to allow  
XX specific binding; and detecting specific binding between each cDNA and at  
XX least one molecule or compound. The molecules or compounds are regulatory  
XX proteins. The combination is useful for preparing a composition for  
XX treating liver disorders associated with steroid therapy, e.g., cirrhosis  
XX or hepatitis. The present sequence represents a human cDNA which is  
XX differentially expressed in steroid-induced C3A liver cells. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from USPTO  
XX at seqdata.uspto.gov/sequence.html.  
XX  
XX Sequence 3745 BP; 1066 A; 755 C; 745 G; 1179 T; 0 U; 0 Other;  
XX  
XX Query Match 83.9%; Score 639.2; DB 12; Length 3745;  
XX Best Local Similarity 90.3%; Pred. No. 2.5e-188;  
XX Matches 699; Conservative 0; Mismatch 63; Indels 12; Gaps 1;  
XX  
XX 1 ATGTCGGGGGGGAGACTTCGGGAATCCGCTGAGGAATTCAGCTGGTGTCTCTGGG 60  
XX 494 ATGTCACCGGGGGAGACTTCGGGAATCCGCTGAGGAATTCAGCTGGTGTCTCTGGG 553  
XX  
XX 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACAGATTTCAGGTATGACAGATTTCGACAAAC 120  
XX 554 GAGCAAAAGCGTTGGAAGACATCTTTGATCACAGATTTCAGGTATGACAGATTTCGACAAAC 613  
XX 121 ACCTATCAGGCAATTAATGGCAATTCAGCTTTTATCAAAACATGATGATGAGGATGGA 180  
XX 614 ACCTATCAGGCAATTAATGGCAATTCAGCTTTTATCAAAACATGATGATGAGGATGGA 673  
XX 181 ACAATCGGCTTCGGGTGAGGATACGGGGGTGAGGAACGTCCTCGTAGCCCTCATTTCC 240  
XX 674 ACAGTACGATTGCAATTTATGGGACACAGCAGGTCAGAGCGGTTTCAGGAGCTTGATTCCT 733

QY 241 AGGTACATCCGTGATTTCTGCTGCAGCTGTAGTAGTTTACGATATCAGAAATGTTAACTCA 300  
DB |||||  
DB 734 AGCTACATTCGTGACTCCACTGTGCGAGTGTGTTTATGATATCAGAAATGTTAACTCA 793  
QY 301 TTCAGCAAACTACAAAGTGGATTGATGATGTACAGACAGAAAGAGGAGTGTATGTTATC 360  
DB |||||  
DB 794 TTCAGCAAACTACAAAGTGGATTGATGATGTACAGACAGAAAGAGGAGTGTATGTTATC 853  
QY 361 ATCAGCTAGTAGGAATAGAAACAGATCTTCTGCAAGAGGCAAGTGTCTAGTTTCAGGAG 420  
DB |||||  
DB 854 ATCATGCTAGTAGGAATAGAAACAGATCTTCTGCAAGAGGCAAGTGTCTAGTTTCAGGAG 913  
QY 421 GGAGAGAGAAAGCCAAAGGGCTGAATGTTTACGTTTATTTGAAACTAGGGCAAAAACCTGGA 480  
DB |||||  
DB 914 GGAGAGAGAAAGCCAAAGAGCTGAATGTTTATTTGAAACTAGTTGCAAAAAGCTGGA 973  
QY 481 TACATGTAAGAGAGCTCTTTTCGAGGTTGACAGCAGCTTTTCCGGGGAATGGAAGACACA 540  
DB |||||  
DB 974 TACAATGTAAAGCAGCTCTTTTCGAGCTGTAGCAGCAGCTTTTCCGGGGAATGGAAGACACA 1033  
QY 541 CAGGACGGAAGCAGAGAGACATGATGACATATAAACTGGAAGAGCCCTCAGGAGCAAAACA 600  
DB |||||  
DB 1034 CAGGACGGAAGCAGAGAGATGATGATGACATATAAACTGGAAGAGCCCTCAGGAGCAAAACA 1093  
QY 601 GTACAGCAAGGGGGTGTGTTCTGCTACTCTCCCATGTCTCTCAACCTT----- 650  
DB |||||  
DB 1094 GTCAAGGAGGAGCTGTCTGCTGCTAACTCCCATGTCTCTCAACCTTCTTCAAGAGC 1153  
QY 651 --TCTCTCAGAAAGCCCTTACTCTTTTCAATGACAGTGTGAATATTTGGCTTGAACCTT 708  
DB |||||  
DB 1154 TCACTGCTTTGGCCCTTACTCTTTCAATGACAGTGTGAATATTTGGCTTGAACCTT 1213  
QY 709 TTCCCTTCATTAATAAGCTTTTGGCAATTCATCATTCGCTGCTGCTGAGGA 762  
DB |||||  
DB 1214 TTCCCTTCAGTAATAAGCTTTTGGCAATTCATCATTCGCTGCTGCTGAGGA 1267  
RESULT 9  
AAFI5989  
ID AAFI5989 standard; cDNA; 3118 BP.  
XX  
XX AAFI5989;  
XX  
XX 13-MAR-2001 (first entry)  
XX  
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:424.  
XX  
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
XX neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
XX vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;  
XX antibacterial; gene therapy; neural; immune; reproductive; renal;  
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
XX wound; infectious disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2000055174-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US005988.  
XX  
XX 12-MAR-1999; 99US-0124270P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI; 2000-587513/55.  
XX P-PSDB; AAB56786.  
XX  
XX Prostate cancer associated gene sequences, referred to as prostate cancer





QY 481 TACAATGTAAGCAGCTCTTCGACGTGTAGCAGCAGCTTTCCGGGAATGGAAGCACA 540  
Db |||||  
QY 481 TACAATGTAAGCAGCTCTTCGACGTGTAGCAGCAGCTTTCCGGGAATGGAAGCACA 540  
Db |||||  
QY 541 CAGGACGGAAGCAGAGACATGAGTGACATATAAACTGGAAAGCCTCAGGAGCAACA 600  
Db |||||  
QY 541 CAGGACGGAAGCAGAGAGATGATGACATATAAACTGGAAAGCCTCAGGAGCAACA 600  
Db |||||  
QY 601 GTCAGGGAAGGGGTGTTCTCTGC 624  
Db |||||  
QY 601 GTCAGTGAAGGAGGTGTTCTCTGC 624  
Db |||||

## RESULT 12

AAZ93835  
ID AAZ93835 standard; DNA; 718 BP.

XX AC AAZ93835;

XX DT 29-AUG-2000 (first entry)

XX DE RAB6 coding sequence.

XX KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
XX antibody; immunogen; mutation; detection; therapy; human; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX CDS 49..675

XX FT /\*tag= a

XX FT /product= "RAB6"

XX PN WO200029625-A1.

XX PD 25-MAY-2000.

XX PF 18-NOV-1999; 99WO-US027630.

XX PR 18-NOV-1998; 98US-0108994P.

XX PA (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

XX PI Duffy HX, Shan J, Yuan L, Budman D, Calabro A;

XX DR WPI; 2000-387828/33.

XX DR P-PSDB; AAY83399.

XX PT Novel nucleic acids encoding wth3 protein useful for inhibiting multiple  
XX drug resistance in tumor cells for identifying the modulators of drug  
XX resistance.

XX PS Example 1; Fig 1a-1b; 82pp; English.

XX CC Nucleic acids encoding WTH3 and RAB6 were isolated by Methy- (or mutant)  
XX Differential Display (MDD) methods and can be used in methods for  
XX detecting methylation patterns in multiple drug resistance loci. Genes  
XX are frequently not methylated in cells where they are expressed but are  
XX methylated in cell types where they are not expressed. Tumour cell DNA is  
XX often methylated to a different extent and in different regions when  
XX compared to DNA of normal cells. The methylation pattern in a multiple  
XX drug resistance locus can be altered and give rise to altered expression  
XX patterns of that multiple drug resistance locus. Nucleic acids  
XX corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
XX can be used as probes for detecting mutations and methylation patterns of  
XX those loci. The nucleic acids and their homologues are useful for  
XX inhibition of multiple drug resistance and for treating tumors exhibiting  
XX multiple drug resistance. They are also useful for detecting and  
XX measuring the expression of mRNA from identified genes and for  
XX determining suitable therapeutic treatment. Antibodies directed against  
XX immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
XX proteins and polypeptides in tissues or body fluids of patients

SQ Sequence 718 BP; 219 A; 141 C; 184 G; 174 T; 0 U; 0 Other;  
Query Match 73.4%; Score 559.6; DB 3; Length 718;  
Best Local Similarity 91.0%; Pred. No. 8e-164;  
Matches 606; Conservative 0; Mismatches 59; Indels 1; Gaps 1;  
QY 1 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTCTTCTCTGGG 60  
Db ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTCTTCTCTGGG 108  
QY 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTGACAAC 120  
Db GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTGACAAC 168  
QY 121 ACCTATCAGGCAATAATTGGCAATTGACCTTTTATCAAAACTATGTACTTCGGAGATGGA 180  
Db ACCTATCAGGCAACAATTTGGCAATTGACCTTTTATCAAAACTATGTACTTCGGAGATGGA 228  
QY 181 ACAATCGGCTTCGGCTGTGGGATACGGCGGTTCAGGAAGCTCTCCGTAGCTCATTTCCC 240  
Db ACAGTACGATTGCAATTTATGGGACACAGCAGGTCAAGAGCGGTTTCAGAGCTTGATTCCT 288  
QY 241 AGGTACATCCGCTGATTCCTGTCAGCTGTAGTGTACGATATCAAAATGTTAACTCA 300  
Db AGCTACATTCGTGACTCCACTGTGGCAGTTGTTGTTATGATATCAAAATGTTAACTCA 348  
QY 301 TTCCAGCAAACTACAAAGTGGATTGATGTTCAGAACAGAAAGAGGAGTGATGTTATC 360  
Db TTCCAGCAAACTACAAAGTGGATTGATGTTCAGAACAGAAAGAGGAGTGATGTTATC 408  
QY 361 ATCAGCGTAGTAGAAATAGAACAGATCTTCTGACAGAGGCAAGTGTCAAGTTGAGGAG 420  
Db ATCATGCTAGTAGGAAATATAAAGCAGATCTTCTGACAGAGGCAAGTGTCAAGTTGAGGAG 468  
QY 421 GGAGAGAGAAAGCAAAAGGCTGAATGTTACGTTTATGAAAAGTAGGGCAAAAGCTGGA 480  
Db GGAGAGAGAAAGCAAAAGGCTGAATGTTACGTTTATGAAAAGTAGGGCAAAAGCTGGA 528  
QY 481 TACAATGTAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGAAATGGAAGCACA 540  
Db TACAATGTAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGAAATGGAAGCACA 588  
QY 541 CAGGACGGAAGCAGAGAGACATCAGTGACATATAAACTGGAAAGCCTCAGGAGCAACA 600  
Db CAGGACGGAAGCAGAGAGAGATATGATTGACATATAAACTGGAAAGCCTCAGGAGCAACA 648  
QY 601 GTCAGCAAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCCCTTCTCTCAGAAG 660  
Db GTCAGTGAAGGAGCTGTTCTCTGCTAATGTCCTTAGTCACTTCAA-CCITTCCTCAGAAG 707  
QY 661 CCCCCT 666  
Db 708 CTCAC 713

## RESULT 13

AAZ93837  
ID AAZ93837 standard; DNA; 740 BP.

XX AC AAZ93837;

XX DT 29-AUG-2000 (first entry)

XX DE RAB6 coding sequence.

XX KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
XX antibody; immunogen; mutation; detection; therapy; human; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 71..697

XX FT /\*tag= a





CC patterns of that multiple drug resistance locus. Nucleic acids  
CC corresponding to the identified loci such as WH3 and RAB6 nucleic acids  
CC can be used as probes for detecting mutations and methylation patterns of  
CC those loci. The nucleic acids and their homologues are useful for  
CC inhibition of multiple drug resistance and for treating tumors exhibiting  
CC multiple drug resistance. They are also useful for detecting and  
CC measuring the expression of mRNA from identified genes and for  
CC determining suitable therapeutic treatment. Antibodies directed against  
CC immunogenic fragments of WH3 and RAB6 are useful for detecting specific  
CC proteins and polypeptides in tissues or body fluids of patients  
XX  
SQ Sequence 573 BP; 162 A; 115 C; 159 G; 137 T; 0 U; 0 Other;

Query Match 69.0%; Score 526; DB 3; Length 573;  
Best Local Similarity 100.0%; Pred. No. 2.2e-133;  
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60  
Db |||||||  
QY 48 ATGTCGGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 107  
QY 61 GAGCAAGCGTTGCAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGACAAC 120  
Db |||||||  
QY 108 GAGCAAGCGTTGCAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGACAAC 167  
QY 121 ACCTATCAGGCAATTAATGGCAATTCACCTTTTATCAAAACTATGATCTGGAGATGGA 180  
Db |||||||  
QY 168 ACCTATCAGGCAATTAATGGCAATTCACCTTTTATCAAAACTATGATCTGGAGATGGA 227  
QY 181 ACAATCGGGCTTCGGCTGTGGGATACGGGGTCAAGAACGCTCTCCGTAGCTCATTTCCC 240  
Db |||||||  
QY 228 ACAATCGGGCTTCGGCTGTGGGATACGGGGTCAAGAACGCTCTCCGTAGCTCATTTCCC 287  
QY 241 AGGTACATCCGTGATCTGCTGCAGCTGTAGTATGATTTACGATATCACAAATGTTAACTCA 300  
Db |||||||  
QY 288 AGGTACATCCGTGATCTGCTGCAGCTGTAGTATGATTTACGATATCACAAATGTTAACTCA 347  
QY 301 TTCCAGCAAACTACAAAGGATTCATGATGTCAAGCAAGAGGAGGATGATGTTATC 360  
Db |||||||  
QY 348 TTCCAGCAAACTACAAAGGATTCATGATGTCAAGCAAGAGGAGGATGATGTTATC 407  
QY 361 ATCAGCTAGTAGGAAATAGAACAGATCTTGCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420  
Db |||||||  
QY 408 ATCAGCTAGTAGGAAATAGAACAGATCTTGCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 467  
QY 421 GGAGAGGAAAGCCAAAGGGCTGAATGTTAGCTTTATGAAACTAGGGCAAAACTGGA 480  
Db |||||||  
QY 468 GGAGAGGAAAGCCAAAGGGCTGAATGTTAGCTTTATGAAACTAGGGCAAAACTGGA 527  
QY 481 TACATGTAAGGACGCTCTTTTCGAGCTGACGAGCTTTGCGCG 526  
Db |||||||  
QY 528 TACATGTAAGGACGCTCTTTTCGAGCTGACGAGCTTTGCGCG 573

RESULT 15  
AAS87692  
ID AAS87692 standard; cDNA; 3195 BP.  
AC AAS87692;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #23496.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX

PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG23505.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 23496; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (I). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3195 BP; 878 A; 698 C; 699 G; 920 T; 0 U; 0 Other;

Query Match 64.3%; Score 489.8; DB 5; Length 3195;  
Best Local Similarity 88.1%; Pred. No. 1.1e-141;  
Matches 700; Conservative 0; Mismatches 62; Indels 33; Gaps 14;  
QY 1 ATGTCGCCGGCGGAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGGTGTCTCTGGGG 60  
Db |||||||  
QY 443 ATGTCGCCGGCGGAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGGTGTCTCTGGGG 502  
QY 61 GAGCAAAAGCGTTG-CAAAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTGACAA 119  
Db |||||||  
QY 503 GAGCAAAAGCTCTGTGAAAGACATCTTTGATCACCAGATTGATGATGACAGTTTGACAA 562  
QY 120 CACCTATCAGGCAATATTCGCATTTGACTTTTATCAAAACTATGATCTGGAGGATGG 179  
Db |||||||  
QY 563 CACCTATCAGGCAACAAATGGCATTTGACTTTTATCAAAACTATGATCTGGAGGATCG 622  
QY 180 AACAAATCGGGCTTCGGCTGTGGGATACGGGGGTGAGGAACTGCTCCGTAGCCCTCATTTCC 239  
Db |||||||  
QY 623 AACAAATCAGGCTTCAGCTGTGGGATACGGGGGTGAGGAACTGCTCCGTAGCCCTCATTTCC 682  
QY 240 CAGTATCAT-CCGTGATCTGCTGCAGCTGTAGTATGATTTAGGATATCACAAATGTTAACT 298  
Db |||||||  
QY 683 CAGTTATATCCCGTGATTTCTGCTGCAGCTGTAGTATGATTTAGGATATCACAAATGTTAACT 742  
QY 299 CATTTCCAGCAAA---CTACAAAGTGGATTGATGATGTCAGAACAGAAAGA-GGAAGTGAT 354  
Db |||||||  
QY 743 CATTTCCAGGCAAACTACAAAGTGGGATTTGATGATGTCAAGCAAGAGGAGGATGAT 802  
QY 355 GTTA-TCATCAGCTAGTAGGAAATAGAACAGATCTTTGCTGACAAAGGCAAGTGTGAGT 413  
Db |||||||



Db 803 GTTATTTCATCATGTCTAGTAGGAAATAAAACAGATCTTGCTGACAAGAGGCAAGTGTCAAT 862  
QY 414 TCAGGAGGAGAGAGAGAGAGCC-AAAGGCTGAATGTT-ACGTTTATTGAAACTAGGCA 471  
Db 863 TGAAGAGGAGAGAGAGAGAGCCAAAGAGCTGAATGTTAATTAATGGAACACTAGTGCA 922  
QY 472 AAAACTGG-ATACAATGTAAAGCAGC-TCTTTGAGGTGTA--GCAGCAGCTTTGCCGGG 527  
Db 923 AAAGCTGGAATACAATGTAAAGCAGCTCTTTGAGCTGTAAGCAGCCAGCTTTGCCGGG 982  
QY 528 AATGAAAGCACACAGGACGGAAGCAGAGACATGAGTGACATAAAACTGGAAGGCC 587  
Db 983 AATGAAAGCACACAGGACGGAAGCAGAGACATATGATTGACATAAACTGGAAGGCC 1042  
QY 588 TCAGGAG-CAAAACAGTCAGCGAAGGGG--TTGTTCTCTGCTACTCTCCCATGTCATCTTC 644  
Db 1043 TCAGGAGCCNACCAGTCAGTGAGGAGGCGCTGTTCTGACTAATCTCCCATGTCATCTTC 1102  
QY 645 AACCTTCTCTCAGAA-----GCCCCCTTACTCTTTCAATTGACTGCAGTG 688  
Db 1103 AACCTTCTCTTTCAGAAAGCTCACTGGCTTTTGGCCCCCTTACTCTTTCAATTGACTGCAGTG 1162  
QY 689 TGAATATTGGCTTGA-ACCTTTTCCCTTCATTAAACGTTTGGCAATTCATCATCTGCTG 747  
Db 1163 TGAATATTGGCTTGAATGCGCTTTTCCCTTCAGTAATAACGTAATTGGGATTCATCATCTGCTG 1222  
QY 748 CCTGTCTCTGTGGAGA 762  
Db 1223 CCTGTCTCTGTGGGA 1237

Search completed: April 25, 2005, 09:44:22  
Job time : 483 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 07:43:50 ; Search time 177 Seconds  
(without alignments)  
7044.310 Million cell updates/sec

Title: US-09-441-857-11  
Perfect score: 762  
Sequence: 1 atgtccgcggcgagactt.....tgcgtcgtctcgtggaga 762

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639.2	83.9	3745	4	US-09-976-594-387
2	564.4	74.1	739	4	US-09-949-016-2946
3	204	26.8	329	4	US-09-513-999C-2679
4	152.4	20.0	86857	4	US-09-949-016-14688
5	135.6	17.8	414	4	US-09-248-796A-6213
6	117.8	15.5	1631	4	US-09-620-312D-587
7	116.2	15.2	719	4	US-09-949-016-3023
8	116.2	15.2	1984	4	US-09-023-655-7
9	111.8	14.7	921	4	US-09-016-434-1124
10	111	14.6	1546	4	US-09-949-016-5540
11	105.6	13.9	24257	4	US-09-949-016-13902
12	102.2	13.4	848	3	US-08-741-411-2
13	101.4	13.3	1255	4	US-09-949-016-1772
14	101.4	13.3	8137	4	US-09-566-921-7
15	99.8	13.1	723	4	US-09-016-434-1422
16	99.8	13.1	920	4	US-09-949-016-4287
17	99.4	13.0	833	4	US-09-620-312D-426
18	98.8	13.0	639	3	US-09-399-913-66
19	98.8	13.0	639	4	US-09-350-614-66
20	98.8	13.0	1148	4	US-09-949-016-4879
21	94	12.3	1775	4	US-09-949-016-4926
22	92.8	12.2	1069	4	US-09-620-312D-646
23	92	12.1	642	4	US-09-248-796A-6190
24	91.8	12.0	1308	4	US-09-270-767-12890
25	90.8	11.9	1090	4	US-09-799-451-914
26	90.6	11.9	601	4	US-09-949-016-73802
27	90.6	11.9	1630	4	US-09-949-016-409

28	90.6	11.9	2874	4	US-09-949-016-2160	Sequence 2160, Ap
29	90.2	11.8	601	4	US-09-949-016-105764	Sequence 105764,
30	89.8	11.8	925	2	US-08-916-901-4	Sequence 4, Appli
31	89.8	11.8	925	3	US-09-154-602-4	Sequence 4, Appli
32	89.6	11.8	875	3	US-09-075-454-10	Sequence 10, Appl
33	89.6	11.8	1106	4	US-09-620-312D-959	Sequence 959, App
34	89.6	11.8	1511	4	US-09-949-016-1134	Sequence 1134, Ap
35	89.6	11.8	1511	4	US-09-949-016-2129	Sequence 2129, Ap
36	89.6	11.8	2612	3	US-09-484-970B-142	Sequence 142, App
37	85.2	11.2	1275	4	US-09-949-016-3508	Sequence 3508, Ap
38	85	11.2	834	4	US-09-943-016-3883	Sequence 3883, Ap
39	83.8	11.0	809	4	US-09-949-016-3376	Sequence 3376, Ap
40	82.6	10.8	1340	2	US-08-824-873-2	Sequence 2, Appli
41	82.6	10.8	1340	3	US-09-198-184-6	Sequence 2, Appli
42	81.8	10.7	842	4	US-09-255-920A-6	Sequence 6, Appli
43	81.6	10.7	998	4	US-09-270-767-1882	Sequence 1882, Ap
44	81.6	10.7	998	4	US-09-270-767-17164	Sequence 17164, A
45	80	10.5	412	4	US-09-513-999C-1698	Sequence 1698, Ap

ALIGNMENTS

RESULT 1  
US-09-976-594-387  
; Sequence 387, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIORITY FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 387  
; LENGTH: 3745  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 107569.15  
US-09-976-594-387

Query Match	83.9%	Score	639.2	DB 4	Length	3745
Best Local Similarity	90.3%	Pred. NO.	7.2e-202			
Matches	699	Conservative	0	Mismatches	63	Indels 12; Gaps 1;
QY	1	ATGTCGGGGGGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG	60			
DB	494	ATGTCACGGGGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG	553			
QY	61	GAGCAACGGTTGCAAGACATCTTTGATCACCAGATTGAGTATGACAGTTTTCACAAAC	120			
DB	554	GAGCAACGGTTGGAAGACATCTTTGATCACCAGATTGATGATGACAGTTTTCACAAAC	613			
QY	121	ACCTATCAGGCAATAATGGCAATGACATTTTATCAAAACTATGTACTTGGAGGATGGA	180			
DB	614	ACCTATCAGGCAATAATGGCAATGACATTTTATCAAAACTATGTACTTGGAGGATGGA	673			
QY	181	ACAATCGGGCTTCGGCTGTGGGATACGGGGGTGAGGAACGTCCTCCGTAGCCTCATTTCC	240			
DB	674	ACAGTACGATTTGCAATTTATGGGACACACAGCGTCAAGAGCGTTTCAGGAGCTTGATTCCT	733			
QY	241	AGGTACATCCGTGATTCCTGTCAGCTGTAGTATTTACGATATACCAAAATGTTAACTCA	300			
DB	734	AGGTACATCCGTGATTCCTGTCAGCTGTAGTATTTACGATATACCAAAATGTTAACTCA	793			
QY	301	TTCCAGCAAACTCAAAAGTGGATTCGATGATCTCAGAACAGAAAGGAGGTGATGTTATC	360			

Db 794 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGAGGAAGTGTATTATC 853  
Qy 361 ATCAGCTAGTAGAAATAGAACAGATCTTCTGTGACAAAGAGCAAGTGTCAAGTTGAGGAG 420  
Db 854 ATCATGCTAGTAGAAATAAAACAGATCTTCTGTGACAAAGAGCAAGTGTCAATTTGAGGAG 913  
Qy 421 GGAGAGGAAAGCCAAAGGCTGAATGTTAGTTCAGTTTATTGAAACTAGGCGCAAAACTGGA 480  
Db 914 GGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTGTGCAAAAGCTGGA 973  
Qy 481 TACAATGTAAAGCAGCTCTTTTCGAGCTGTAGCAGCAGCTTTTCGCGGGAATGAAAGACACA 540  
Db 974 TACAATGTAAAGCAGCTCTTTTCGAGCTGTAGCAGCAGCTTTTCGCGGGAATGAAAGACACA 1033  
Qy 541 CAGACGGAAGCAGAGAAAGATGATGATGACATATAAACTGGAAAAAGCCCTCAGGAGCAAAACA 600  
Db 1034 CAGGACAGAAAGCAGAGAAAGATGATGATGACATATAAACTGGAAAAAGCCCTCAGGAGCAAAACA 1093  
Qy 601 GTCAGCGAAGGGGTGTTCTGCTGCTACTCTCCCATGTCATCTTCAACCT----- 650  
Db 1094 GTCAGTGAAGGAGGTGTTCTGCTGCTTAATCTCCCATGTCATCTTCAACCTTTTCAGAAGC 1153  
Qy 651 --TCTCAGAAAGCCCTTACTCTTTCAATGACTGCAGTGTGAATATTGGCTTGAACCTT 708  
Db 1154 TCACTGCTTTGCCCCCTTACTCTTTCAATGACTGCAGTGTGAATATTGGCTTGAACCTT 1213  
Qy 709 TTCCCTTCATTAATAACCTTTTGGCAATTCATCATCTGCTGCTGCTCTCTGTGAGGA 762  
Db 1214 TTCCCTTCAGTAATAACCTTTTGGCAATTCATCATCTGCTGCTGCTGCTGAGGA 1267

RESULT 2  
US-09-949-016-2946  
; Sequence 2946, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2946  
; LENGTH: 739  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-2946

Query Match 74.1%; Score 564.4; DB 4; Length 739;  
Best Local Similarity 91.4%; Pred. No. 2.2e-177;  
Matches 609; Conservative 0; Mismatches 56; Indels 1; Gaps 1;  
Qy 1 ATGTCGCGGGCGGAGACTTCGGGAAATCCGCTGAGGAAATTCAGAGCTGGTTCCTCGGG 60  
Db 70 ATGTCGCGGGCGGAGACTTCGGGAAATCCGCTGAGGAAATTCAGAGCTGGTTCCTCGGG 129  
Qy 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCAGAAC 120  
Db 130 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCAGAAC 189  
Qy 121 ACCTATCAGGCAATTAATTTGGCAATTCAGTTTATCAAAAATATGATGATCTGGAGGATGGA 180  
Db 190 ACCTATCAGGCAAAATTTGGCAATTCAGTTTATCAAAAATATGATGATCTGGAGGATGGA 249  
Qy 181 ACAATCGGGCTTCGGCTGTGGATACGCGGGTCAGGAAACGCTCTCCGTAGCCTCATTTCCC 240

Db 250 ACAGTACGATTGCAATTTATGGGACACAGCAGGTTCAAGCGGTTTCAGGAGCTTGAATCCT 309  
Qy 241 AGGTACATCCGTTGATTCGTCGAGCTGTAGTAGTTTACGATATCAAAATGTTAACTCA 300  
Db 310 AGCTACATTCGTTGACTCCACTGTGGCAGTTGTTGTTTATGATATCAGAAATGTTAACTCA 369  
Qy 301 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGAGGAAGTGTATCTATC 360  
Db 370 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGAGGAAGTGTATCTATC 429  
Qy 361 ATCAGCTAGTGTAGAAATAGAACAGATCTTCTGACAAAGAGGCAAGTGTCAAGTTGAGGAG 420  
Db 430 ATCATGCTAGTGTAGAAATAGAACAGATCTTCTGACAAAGAGGCAAGTGTCAATTTGAGGAG 489  
Qy 421 GGAGAGAGGAAAGCCAAAGGCTGAAATGTTTACGTTTATGAAACTAGGCGCAAAACTGGA 480  
Db 490 GGAGAGAGGAAAGCCAAAGAGCTGAAATGTTTATGTTTATGAAACTAGTGTCAAAAGCTGGA 549  
Qy 481 TACAATGTAAAGCAGCTCTTTTCGACGTTAGCAGCAGCTTTTCGCGGAAATGGAAGACACA 540  
Db 550 TACAATGTAAAGCAGCTCTTTTCGACGTTAGCAGCAGCTTTTCGCGGAAATGGAAGACACA 609  
Qy 541 CAGGACGGAAGCAGAGAAAGACATGAGTGACATATAAACTGGAAAAAGCCCTCAGGAGCAAAACA 600  
Db 610 CAGGACGGAAGCAGAGAAAGATATGATGACATATAAACTGGAAAAAGCCCTCAGGAGCAACCA 669  
Qy 601 GTCAGCGAAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCTTCTCTCAGGAG 660  
Db 670 GTCAGTGAAGGAGCTGTTCTCTGCTAATCTCTCCCATGTCATCTTCAA--CCTTCTTCAAGAG 728  
Qy 661 CCCCCT 666  
Db 729 CTCAC 734

RESULT 3  
US-09-513-999C-2679  
; Sequence 2679, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 2679  
; LENGTH: 329  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 99..329  
US-09-513-999C-2679

Query Match 26.8%; Score 204; DB 4; Length 329;  
Best Local Similarity 77.8%; Pred. No. 1.7e-57;  
Matches 246; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
Qy 11 GCGGAGACTTCGGGAATCCGCTCAGGAAATTCAGAGCTGGTGTCTCTCGGGGAGCAAAAGCG 70  
Db 10 GGGAGATTTTGGGAATCCACTGAGAAATTCAGTTGGTGTCTCTCGGGGAGCAAGCG 69  
Qy 71 TTGCAAGAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTGTGACAAACCTTATCAGG 130  
Db 70 TCGGGAAGACGTCCTCTGATTACGAGGTTTCATGTACGACAGCTTCGACAAACATATCCAGG 129



; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt\_FL\_genes Version 1.0
; SEQ ID NO 587
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (66)..(650)
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: (1)..(1631)
; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-587

Query Match 15.5%; Score 117.8; DB 4; Length 1631;  
Best Local Similarity 53.1%; Pred. No. 2.6e-28;  
Matches 251; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

Qy	29	CGCTGAGGAATTC	CAAGCTGGTTCCTGGGGGACCAAGCGTTGC	CAAGACATCTTGA	88
Db	70	CGCTGAGGAGCT	CAAGTGTCTGCTCGGGATACAGGTGTAGTAAATCGGATATG	129	
Qy	89	TCACGAGATT	CAGGTATGACAGTTTTCGAGCAACACCTATCAGGCAATAATTTGGCATTGACT	148	
Db	130	TGTGCGGTTT	TGTGGAAGACAGTTTTTGTATCAAAACATCAACCCAAATAGGGGCATCTT	189	
Qy	149	TTTTATCAAA	ACTATGTACTTGGAGGATGGAAACATCGGGCTTCGGCTTGGGGATACGG	208	
Db	190	TTATGACCA	AGACTGTCAGTACCAAAATGAGCTACATAAAATTCCTAATCTGGGATACAG	249	
Qy	209	CGGTCACGA	AACTGCTCGTAGCTCATTCCAGGTACATCGTGATTCTGTCGAGCTG	268	
Db	250	CTGGAACA	AGACGATTTCTGCTTAGCACCAGTATCTATCGAGGTCGGCTGCAGCTA	309	
Qy	269	TAGTAGTTT	ACGATATCAAAATGTTAACTCATTCACGCAAACTACAAAGTGGATTGATG	328	
Db	310	TAATCGTT	TATGATATCAAAAGAAGAGACATTTTCAACATTAAGAATTTGGGTGAAAG	369	
Qy	329	ATGTCAGA	ACAGAAAGAGTGTATGATCATCACCTAGTAGGAATAAGACAGATC	388	
Db	370	AGCTTCGA	CAGCATGCGCCACCTAATATTGTAGTTGCCATTGCAGGAAATAAATGTGATC	429	
Qy	389	TTGCTGCA	AGAGCGCAAGTGTGAGTTGAGGGGAGAGGAAAGCCAAAGGGCTGAATG	448	
Db	430	TTATCGAT	GTAAAGAGTGTGAGAGAGATGCAAGGACTACGCCGACTCTATTCTATG	489	
Qy	449	TTACGTTT	TATTGAAACTAGGGCAAAACTGGATACAATGTAAAGCAGCTCTTT	501	
Db	490	CAATTTT	TGTAGAGACCGCGCAAAACCGGATAAATCAATGAACTCTTT	542	

RESULT 7  
US-09-949-016-3023  
; Sequence 3023, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3023  
; LENGTH: 719  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-3023

Query Match 15.2%; Score 116.2; DB 4; Length 719;  
Best Local Similarity 51.7%; Pred. No. 5.2e-28;  
Matches 265; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

Qy	22	GGGAATCCG	TGAGGAATTC	CAAGCTGGTTCCTGGGGGAGCAAGCGTTGCAAGACA	81
Db	99	GGAAATAAA	TATATGCCAGTTC	CAAACTAGTACTTCTGGGAGAGTCCGCTGTTGGCAATCA	158
Qy	82	TCCTTTGAT	CACCAATTC	CAGGTATGACAGTTTTTCACAAACACCTATCAGGCAATAATTTGGC	141
Db	159	AGCCTAGT	GGTTCGTTTT	TGTGAAAGGCCAATTCATGAATTTCAAGAGAGTACCATTTGGG	218
Qy	142	ATTGACTTT	TATCAAAAAC	TATGTACTTGGAGGATGGAACAATCGGGCTTCGGCTGTGG	201
Db	219	GCTGCTTT	TCTTAACCCAA	ACTGTATGTTGTATGACACTACAGTAAAGTTTGAATAATGG	278
Qy	202	GATACGGG	GGGTACGA	ACGTCCTCCGTAGCCTCATTCAGGTCATCCCGTATCTTGCT	261
Db	279	GATACAGC	TGGTCA	GAAGACGATACCATAGCCTACGCCAATGTACTACAGAGGAGCAAA	338
Qy	262	GCAGCTGT	AGTATGAT	TATCACAATGTTAACTCATTCAGCAAACTACAAAAGTGG	321
Db	339	GCAGCCAT	GTATGATAT	GATATACAAATGAGAGTCCCTTTGCAAGAGCAAAAATTTGG	398
Qy	322	ATTGATCAT	GTACAG	ACAGAAAGGAAGTGAATGTTATCATCACGCTAGTAGGAAATAGA	381
Db	399	GTTAAAGA	AACTTCAG	AGGCAAGCAAGTCTTAACATTTGTAATAGCTTTTATCGGAAACAAAG	458
Qy	382	ACAGATCT	TCTGAC	AGAGGCAAGTCTCAGTTTGAGGGGAGAGAGGAAGCCAAAGGG	441
Db	459	GCCAGCCT	AGCAATA	AAAGAGCAGTAGATTTCCAGGAAGCACAGTCTCTATGCAGATGAC	518
Qy	442	CTGAATCT	GTACGTTT	TATGAAACTTAGGGCAAAACTGGATACAAATGTAAAGCAGCTCTTT	501
Db	519	AATAGTAT	TATTTAT	TATGAGACATCCGCTTAAACATCAATGATGTAAATGAATATTC	578
Qy	502	CGAGCTGT	AGCAGC	AGCTTTGCCGGGAATGGA	534
Db	579	ATGGCAAT	AGTAGTAA	AAAAATTTGCCAAAGATGAA	611

RESULT 8  
US-09-023-655-7  
; Sequence 7, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: U937NOT01  
CLONE: 000513  
US-09-023-655-7

Query Match 15.2%; Score 116.2; DB 4; Length 1984;  
Best Local Similarity 51.7%; Pred. No. 1e-27;  
Matches 265; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 22 GGAATCCGCTGAGGAATTCAGCTGGTTCCTGGGGAGCAACGCTTGCAGAAC 81  
DB 11 GGAATAAATATGCCAGTTCAAATAGTACTTCTGGGAGAGTCCGCTGTGCAATCA 70  
QY 82 TCTTTGATCACCAGATTCAGGTATGACAGTTTGACACACCTATCAGGCAATATGGC 141  
DB 71 AGCCTAGTCTCGTTTCTGAAAGGCCAATTCATGAATTTCAAGAGAGTACCATTTGG 130  
QY 142 ATTGACTTTTATCAAAACTATGTACTTGGAGATGGAACAACTCGGCTTCGGCTGTGG 201  
DB 131 GCTGCTTTTCTAACCCAACTGTATGTCTGTATGACACATACAGTAAAGTTGAAATGG 190  
QY 202 GATACGGCGGTGAGGAAGCTCTCCGTAGCCTCAITCCAGGTACATCCGTGATTTGCT 261  
DB 191 GATACAGCTGGTCAAGAACGATACCATAGCTAGCACCACCAATGTACTACAGAGGAGCACA 250  
QY 262 GCAGCTGTAGTATTTACATATACAAATGTTAACTCATTCACGCAAACTCAAAAGTGG 321  
DB 251 GCAGCCATAGTTGTATATGATATACAAATGAGGAGTCTCTTTGCAAGAGCAAAAATTTGG 310  
QY 322 ATTGATGATGTGACAGACAGAGAGAGAGTGTATGTTTATCATCACGCTAGTAGGAATAGA 381  
DB 311 GTTAAAGAACTTCAGAGGCAAGCAAGTCTTAACATTTGTAATAGCTTTTATCGGGAACAG 370  
QY 382 ACAGATCTTCTGACAGAGGCAAGTGTGAGTTGAGGAGGAGAGAGAAAGCRAAGGG 441  
DB 371 GCGACCTAGCAATAAAGAGCAGTATGATTTCCAGAGAGCAGTCTCTATGCAATGAC 430  
QY 442 CTGAATGTTACGTTTATGAACTAGGCAAAAACCTGGATACATGTAAGCAGCTCTTT 501  
DB 431 AATAGTTTATTTATTCATGAGACATCCGCTAAACATCAATGAATGAAATGAAATATTC 490  
QY 502 CGACGTTGAGCAGCAGCTTTCCGGGAATGGAA 534  
DB 491 ATGGCAATAGTAAAAAATTTGCCAAGAAATGAA 523

RESULT 9  
US-09-016-434-1124  
Sequence 1124, Application US/09016434  
Patent No. 6500938  
GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhamer

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,434  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0002 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1124:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: g1388194  
US-09-016-434-1124

Query Match 14.7%; Score 111.8; DB 4; Length 921;  
Best Local Similarity 51.5%; Pred. No. 1.8e-26;  
Matches 257; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

QY 8 CGGGGGGAGACTTCGGGAATCCGCTGAGGAATTCAGCTGGTGTCTCTGGGGGAGCAA 67  
DB 44 CTGGCTCGGAGCACATGATGGGATACGGGAGCTCAAAAGTGTGCTTCTCGGGGACACTG 103  
QY 68 GGGTTGCAAGACATCTTTTGATCACAGATTCAGGTATGACAGTTTTCACACACACTATC 127  
DB 104 GGGTTGGAATCAAGCATCGTGTGATTTGTCAGGATCATTTCACCAACATCA 163  
QY 128 AGGCAATAATTGGCATTCACATTTTATCAAAAACCTATGTACTTTGAGGATGGAACAATGG 187  
DB 164 GCGCTACTATTGGGGCATCTTTTATGACCAAACTGTGCTTGTGCAAAATGAATTCACA 223  
QY 188 GCGTTGCGCTGTGGATACGGGGTCAAGACGCTCCGTAGCCTCATTTCCAGGTACA 247  
DB 224 AGTTCTCATCTGGGACATCGTGTGTCAGGAACGGTTTCATTTCATTGGCTCCCATGTACT 283  
QY 248 TCCGTGATTTGCTGCTGAGTGTAGTTTACGATATCACAAATGTTAACTCATTTCCAGC 307  
DB 284 ATCGAGGCTCAGTCTGACGTGTTATCGTGTATGATATTTACCAGCAGGATTCATTTTATA 343  
QY 308 AAACTACAAAGTGGATTGATGTGACAGAACAGAAAGAGGAGTGTATGTTATCATCACGC 367  
DB 344 CTTTGAAGAAATGGCTCAAGGAGCTGAAAGAACATGTTCCAGAAACATTTGTAATGGCCA 403  
QY 368 TAGTAGGAAATAGAACAGATCTTGTGTCAGAGAGGCAAGTGTGTCAGTTGAGGAGGAGAGA 427  
DB 404 TCGCTGGAACCAAGTGCAGCTCTCAGATATTTAGGGAGGTTCCCTCTGAAGGATGCTAAGG 463

```
Qy 428 CGAAGCCAAAGGGCTGAATGTTACGTTTATTGAACCTAGGCAAAAACCTGGATACAATG 487
Db 464 BATAAGCTGAATCCATAGTGCCTCGTGGTTGAGACAGTGCAAAAAATGCTATTATA 523

Qy 488 TAAAGCAGCTCTTTTCGACG 506
Db 524 TCGAAGAGCTCTTTCAAGG 542

RESULT 10
US-09-949-016-5540
; Sequence 5540, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5540
; LENGTH: 1546
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5540

Query Match 14.6%; Score 111; DB 4; Length 1546;
Best Local Similarity 52.0%; Pred. No. 4.6e-26;
Matches 249; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 22 GGAATCCGCTGAGGAATTCAGCTGTTCTCTGGGGGAGCAAGGTTGCAAGACA 81
Db 134 GGAACAAGATCTGTCAATTTAAAGCTGTTCTCTGGGGGAGTCTCGCGTAGGCAATCC 193

Qy 82 TCTTTGATCACCAGATTCAGGTATCACAGTTTTCACACACCTATCACGGAATATTTGGC 141
Db 194 AGCCTGCTCTCGCTTTGTCAAGGACAGTTTCAAGAGTACAGGAGACCAATTTGGA 253

Qy 142 ATTGACTTTTATCAAAAACCTATGTAATTGAGGATGGAACAATCGGGCTTCGGCTGTGG 201
Db 254 GCGGCTTTCTCCACACAGACTGTCTGCTGGATGACACACAGTCAAGTTTGAGATCTGG 313

Qy 202 GATACGGGGGTGAGGAAGCTCTCGTAGCCTCATTCACAGTATCATCCAGTATCTGCT 261
Db 314 GACACAGCTGACAGGAGCGGTATCACAGCTTGGCCCCCATGTACTATCGGGGGGCCAG 373

Qy 262 GCAGCTGTAGTAGTTTACGATATCAAAATGTTAACTCATTCAGCAAACTACAAAGTGG 321
Db 374 GCTGCCATCGTGCTATGATCATCCAAACAGATATTTGACGCGGCCAAGACTGG 433

Qy 322 ATTGATGATGTGAGAACAGAGGAAAGTATGTTATCATCACGCTAGTAGGAAATAGA 381
Db 434 GTGAAGGAGCTACAGAGGACGAGCCAGCCCCAACATTCGTCATTGCACTCGCGGGTAAACAAG 493

Qy 382 ACAGATCTTGTGACAAAGGCAAGTGTCAAGTTGAGGAGGAGAGAGAAAGCCAAAGGG 441
Db 494 GCAGACCTGGCCAGCAAGAGAGCGGTGGAATTTCCAGGAAGCAGCAAGCCTATGCGAGCGAC 553

Qy 442 CTGAATGTTACGTTTATTGAACCTAGGCAAAAACCTGGATACAATGTAAAGCAGCTCTT 500
Db 554 AACAGTTTGTCTTCATCGACATCAGCAAAAGACTGCAATGAACGTGAACGAATCTT 612

RESULT 11
US-09-949-016-13902/c
```

```
; Sequence 13902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13902
; LENGTH: 24257
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13902

Query Match 13.9%; Score 105.6; DB 4; Length 24257;
Best Local Similarity 51.7%; Pred. No. 1.7e-23;
Matches 240; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 40 TTCAGCTGTGTTCCTGGGGGAGCAAAAGCTTGCAAAAGACATCTTTGATCACCAGATTC 99
Db 8910 TTCAAAGTTGCTGTGATCGGGGACTCGGGGGTGGGCAAGACTTGTCTGATCATTCGCTTT 8851

Qy 100 AGGTATGACAGATTTTGACAAACCTATCAGGCAATATTCGCATTTGACTTTTATCAAAA 159
Db 8950 GCAGAGGACAACTTCAACAAACCTTATCATCTCCACCATCGAATTTGATTTCAAGATCGC 8791

Qy 160 ACTATGTACTTGGAGGATGGAACAATCGGGCTTCGGCTGTGGGATACGGGGGTCAAGAA 219
Db 8790 ACTGTGATATAGAGGGGAAGAGATCAAACTACAAGTCTGGGACACAGCTGGCCAAGAG 8731

Qy 220 CGTCTCGTAGCTTCATCCAGGTATCATCGGTGATTTCTGCTGACGCTGTAGTATTAC 279
Db 8730 CGGTTCAAGACAAATACTACTGCTACTACCGTGGAGCCATGGGCATTATCTTAGCATAC 8671

Qy 280 GATATCAAAATGTTAACTCATTCACGCAAACTACAAGTGGATTTGATGATGTCAGAA 339
Db 8670 GACATCAGGATGAGAAATCTTCGAGAAATTTCAAGACTGGATGAAAGCATCTGAG 8611

Qy 340 GAAAGAGGAAGTGTATTTATCATCACGCTAGTAGGAATAGAACAGATCTTTGCTGCAAG 399
Db 8610 AATGCTCTCAGCTGGGGTGGAGACCTCTTCTAGGGAACAAAATGTGACATGGAGGCCAAG 8551

Qy 400 AGGCAAGTGTCAAGTTGAGGAGGAGAGAGAAAGCCAAAGGCTGAATGTTACGTTTATT 459
Db 8550 AGGAAGTGTGAGAGGAGAGCGGCGGATAGTTGGCTCGAGAGCATGGAATCCGATTTTC 8491

Qy 460 GAAACTAGGCAAAAACCTGGATACAATGTAAGCAGCTCTTTTCG 503
Db 8490 GAAACTAGTCTAAATCCAGTATGAATGTGATGAGGCTTTTAG 8447

RESULT 12
US-08-741-411-2
; Sequence 2, Application US/08741411
; Patent No. 6124116
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: NOVEL RAB PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
```



STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/741,411  
FILING DATE: Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0139 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

## INFORMATION FOR SEQ ID NO: 2:

## SEQUENCE CHARACTERISTICS:

LENGTH: 848 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:

## LIBRARY:

CLONE: Consensus

US-08-741-411-2

Query Match 13.4%; Score 102.2; DB 3; Length 848;

Best Local Similarity 51.7%; Pred. No. 2.7e-23;

Matches 233; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 56 TGGGGGAGCAAGCGTTGCAAGACATCTTTCATCACCAGATTCAGGTATGACAGTTTG 115  
DB 211 TGGAGGACACTGGGTTGGAAATCAAGCATCGTGTGATTTGTCAGGATCACTTTG 270  
QY 116 ACAACACCTATCAGGCAATAATGGCAATGACTTTTATCAAAAACATGATCTGGAGG 175  
DB 271 ACCAGACATCAGCCCTACTATTGGGCATCTTTATGACCAAACTGTGCCTTGGGA 330  
QY 176 ATGGAACTTCGCGCTTCGGCTGTGGGATACGGCGGGTCAGGAACGTCTCCGTAGCCTCA 235  
DB 331 ATGAACCTTCAAAAGTTCCCTCATCTGGGACACTGCTGTCAGGAACGGTTTCATTCTGG 390  
QY 236 TTCCAGGTACATCCGTGATCTGTCGAGCTGTAGTAGTTTACGATATCACAATGTGA 295  
DB 391 CTCCATGTACTATCGAGGCTCAGCTGCAGCTGTTATCGTGTATGATATACCAAGCAGG 450  
QY 296 ACTCATTCCAGCAAACTACAAAGTGGATTGATGTCAGACAGAAAGAGGAAGTGTATG 355  
DB 451 ATTCATTTTATACCTTGAAGAAATGGTCAAGAGCTGAAAGAACTATGTCAGAAACA 510  
QY 356 TTATCATCATCGCTAGTAGGAATAGAACAGATCTTCTGCAAGAGGCAAGTGTCAAGTTG 415  
DB 511 TTGTAATGCCATCGCTGGAACAAGTGGGACCTCTCAGATATTAGGAGGTTCCCTGTA 570  
QY 416 AGAGGAGAGAGGAAGCAAGGCTGAATGTTAGCTTTTATGAACTAGGGCAAAA 475  
DB 571 AGGATGCTAAGGAATACGCTGAATCCATAGGTGCCATCGTGTGAGACAAGTGCAGAAA 630  
QY 476 CTGGATACAAATGTAAGACAGCTCTTTCCAGC 506  
DB 631 ATGCTATTAATTCGAAGAGCTCTTTCAAGG 661

RESULT 13

US-09-949-016-1772

Sequence 1772, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CLO01307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1772

LENGTH: 1255

TYPE: DNA

ORGANISM: Human

US-09-949-016-1772

Query Match 13.3%; Score 101.4; DB 4; Length 1255;

Best Local Similarity 50.7%; Pred. No. 6.4e-23;

Matches 243; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 39 ATTCAAGCTGTGTTCTGGGGGAGCAAGCGTTGCAAGACATCTTTGATCACCAGATT 98  
DB 129 ATTCAAGTTACTTCTGATTGGCGACTCAGGGGTTGGAAAGTCTTGCCCTTCTTTAGGTT 188  
QY 99 CAGGTATGACAGAGTTTGTCAACACCTATCAGGCAATAATTGGCATTGACTTTTATCAAA 158  
DB 189 TGCAGATGATACATATACAGAAAGCTACATCAGACATTTGGTGGATTTCAAAATAAG 248  
QY 159 AACTATGTTACTTGGAGGATGAACAATCGGCTTCGGCTGTGGGATACGGGGGTGAGGA 218  
DB 249 AACTATAGTTAGACGGGAAAAACAATCAAGCTTCMAATATGGGACACAGCAGGCGCAGGA 308  
QY 219 AGCTCCGTAGCTTCATTCCTCAGGTACATCCGTGATTTCTGTCAGCTGTAGTAGTTTA 278  
DB 309 AAGATTTCGAACAATCACCTCCAGTTATTACAGAGGAGGCCATGGCATCATAGTTGTGA 368  
QY 279 CGATATCAAAATGTTAACTCATTCAGCAAACTACAAAGTGGATTGATGTGAGAAC 338  
DB 369 TGATGTGACATCAGGAGTCTTCAATAATGTTAAACAGTGGCTGCAGGAATAGATCG 428  
QY 339 AGAAAGAGGAAGTGTATTCATCACGCTAGTAGGAATAAGAACAGATCTTGTGTGACAA 398  
DB 429 TTATGCCAGTGAAATGTCAACAATTTGTTAGGGAAACAAATGTGATCTGACCACAAA 488  
QY 399 GAGGCAAGTGTGTTGAGGAGGAGAGAGGAAGCAAGGCTGAATGTTACGTTTAT 458  
DB 489 GAAAGTAGTAGACTACACAACAGCGAAGGAATTTGCTGATTTCCCTTGGAAATTCGCTTTT 548  
QY 459 TGAAGTACAGGCAAAACTGGATACAATGTTAAAGCAGCTCTTTCGACGCTGAGCAGCAG 517  
DB 549 GGAACACAGTGTAGAAATGCAACGAATGTAGAACAGTCTTTTCATGACGATGGCAGCTG 607

RESULT 14

US-09-566-921-7

Sequence 7, Application US/09566921

Patent No. 6682888

GENERAL INFORMATION:

APPLICANT: Loring, Jeanne F.

APPLICANT: Tingley, Debora W.

APPLICANT: Edwards, Carla M.

TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE

FILE REFERENCE: PA-0024 US

CURRENT APPLICATION NUMBER: US/09/566,921

CURRENT FILING DATE: 2000-05-05

```
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 8137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6682888 411474.10
; NAME/KEY: unsure
; LOCATION: 3488-3788
; OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-7

Query Match      13.3%; Score 101.4; DB 4; Length 8137;
Best Local Similarity 50.7%; Pred. No. 2.1e-22;
Matches 243; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

Qy 39 ATTCAAGCTGGTTCCTGGGGGAGCAAGCGTTGCAAGACATCTTTGATCACCAGATT 98
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 235 ATTCAAGTTACTTCTGATTGGCGACTCAGGGGTTGGAAGTCTTGCCCTTCTTTAGGTT 294
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 99 CAGGTATGACAGTCTTTGCAACACCTATCAGGCAATAATTGGCAATTGACTTTTATCAAA 158
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 295 TGCAGATGATACATATACAGAAAGCTACATCAGCACAAATTGGTGGATTTCAAAATAAG 354
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 159 AACTATGTACTTGGAGGATGGAACAATCGGCTTGGCTGTGGATACGGCGGTTCAGGA 218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 355 AACTATAGATTAGACGGGAAACAATCAAGCTTCAAAATATGGACACAGCAGGCCAGGA 414
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 219 ACGTCTCGTAGCTCATCTCCAGGTACATCGGTGATTCGTCTGAGCTGTAGTAGTTTA 278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 415 AAGATTTTCCGAACAATCACCTCCAGTTATTACAGAGGAGCCCATGGCATCATAGTTGTGA 474
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 279 CGATATCAAAATGTTAACTCATCTCCAGCAAACTACAAAGTGGATTTGATGATGCAGAAC 338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 475 TGATGTGACATCAGGAGTCTCTCAATAATGTTAAACAGTGGCTGCAGGAATAGATCG 534
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 339 AGAAAGAGGAAGTGTATTCATCACGCTAGTAGGAAATAGAACAGATCTTGCTGACAA 398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 535 TTATGCCAGTGAAATGTCAACAAATTTGTTGGTAGGAAACAAATGTGATCTGCACACAAA 594
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 399 GAGGCAAGTGTGAGTGGAGGAGAGAGAGAAAGCCAAAGGCTGATGTAGTTTAT 458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 595 GAAAGTAGTAGCTACACACAGCGAAGGAATTTGCTGATTCCTTGGAAATTCGGTTTTT 654
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 459 TGAAGTGGGCAAAAATCGATACAAATGTAAGCAGCTCTTTCGACGTGTAGCAGCAG 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 655 GGAACCCAGTCTAAGATGCAACGAATGTAGAACAGTCTTTCATGACGATGGCAGCTG 713
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 15

```
US-09-016-434-1422
; Sequence 1422, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1422:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 9550059
US-09-016-434-1422
```

```
Query Match      13.1%; Score 99.8; DB 4; Length 723;
Best Local Similarity 50.5%; Pred. No. 1.5e-22;
Matches 242; Conservative 0; Mismatches 237; Indels 0; Gaps 0;
```

```
Qy 39 ATTCAAGCTGGTTCCTGGGGGAGCAAGCGTTGCAAGACATCTTTGATCACCAGATT 98
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 83 ATTCAAGTTACTTCTGATTGGCGACTCAGGGGTTGGAAGTCTTGCCCTTCTTTAGGTT 142
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 99 CAGGTATGACAGTCTTTGACAAACCTATCAGGCAATAATTGGCAATTCACATTTTATCAAA 158
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 143 TGCAGATGATACATATACAGAAAGCTACATCAGCACAAATTGGTGGATTTCAAAATAAG 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 159 AACTATGTACTTGGAGGATGGAACAATTCGGCTTGGGATACGGCGGTTCAGGA 218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 203 AACTATAGAGTTAGACGGGAAACAATCAAGCTTCAAAATATGGACACAGCAGGCCAGGA 262
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 219 ACGTCTCGTAGCTCATCTCCAGGTACATCGGTGATTCGTCTGACGCTGTAGTAGTTTA 278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 263 AAGATTCGAACAATCACCTCCAGTTATTACAGAGGAGCCCATGGCATCATAGTTGTGA 322
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 279 CGATATCAAAATGTTAACTCATTTCCAGCAAACTACAAAGTGGATTTGATGATGCAGAAC 338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 323 TGATGTGACAGATCAGGAGTCTCTTCAATAATGTTAAACAGTGGCTGCAGGAATAGATCG 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 339 AGAAAGAGGAAGTGTATTCATCATCGCTAGTAGGAAATAGAACAGATCTTGCTGACAA 398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 383 TTATGCCAGTGAAATGTCAACAAATTTGTTGGTAGGAAACAAATGTGATCTGCACACAAA 442
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 399 GAGGCAAGTGTGAGTGGAGGAGAGAGAAAGCCAAAGGCTGATGTAGTTTAT 458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 443 GAAAGTAGTAGCTACACACAGCGAAGGAATTTGCTGATTCCTTGGAAATTCGGTTTTT 502
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 459 TGAAGTGGGCAAAAATCGATACAAATGTAAGCAGCTCTTTCGACGTGTAGCAGCAG 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 503 GGAACCCAGTCTAAGATGCAACGACGATAGAACAGTCTTTTCATGACGATGGCAGCTG 561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: April 25, 2005, 11:43:33  
Job time : 179 secs



NAME/KEY: CDS  
LOCATION: (27078)..(27094)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (141192)..(141769)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (159571)..(159606)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (174525)..(174575)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (234891)..(235013)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (235514)..(235560)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (279677)..(279729)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (408660)..(409123)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (409204)..(409669)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (428381)..(428396)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (472204)..(472330)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (714252)..(714355)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (714447)..(714529)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (739794)..(739891)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (744484)..(744602)  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (51812)..(51911)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (57122)..(57221)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (79368)..(79467)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (293951)..(294050)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (310089)..(310188)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (332935)..(332935)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (332992)..(332992)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:

NAME/KEY: modified\_base  
LOCATION: (362002)..(362101)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (639781)..(639880)  
OTHER INFORMATION: a, t, c, g, unknown or other  
US-10-292-798-1369  
Query Match 99.7%; Score 759.4; DB 17; Length 744802;  
Best Local Similarity 99.9%; Pred. No. 6.2e-227;  
Matches 760; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTTCCTCGGG 60  
Db 345708 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTTCCTCGGG 345649  
Qy 61 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTGTGACAAC 120  
Db 345648 GAGCAAAAGCGTTGCAAAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTGTGACAAC 345589  
Qy 121 ACCTATCAGCAATAATTGGCATTTGACTTTTATCAAAAACTATGTACTTTGGAGATGGA 180  
Db 345588 ACCTATCAGCAATAATTGGCATTTGACTTTTATCAAAAACTATGTACTTTGGAGATGGA 345529  
Qy 181 ACAATCGGCTTCGGCTGTGGGATACGGCGGCTCAGGAACGTCTCCGTAGCCTCATTTCCC 240  
Db 345528 ACAATCGGCTTCGGCTGTGGGATACGGCGGCTCAGGAACGTCTCCGTAGCCTCATTTCCC 345469  
Qy 241 AGGTACATCCGTGATTCGCTGCAGCTGTAGTATTTTACGATATACAAATGTTAACTCA 300  
Db 345468 AGGTACATCCGTGATTCGCTGCAGCTGTAGTATTTTACGATATACAAATGTTAACTCA 345409  
Qy 301 TTCAGCAAACTACAAAGTCGATTTGATGTTCAGAACAGAAAGAGGATGTATGTTATC 360  
Db 345408 TTCAGCAAACTACAAAGTCGATTTGATGTTCAGAACAGAAAGAGGATGTATGTTATC 345349  
Qy 361 ATCAGCTAGTAGGAATAGACAGATCTTCTGCAGAGGCAAGTGTCAAGTTGAGGAG 420  
Db 345348 ATCAGCTAGTAGGAATAGACAGATCTTCTGCAGAGGCAAGTGTCAAGTTGAGGAG 345289  
Qy 421 GGAGAGAGAAAGCCAAAGGCTGAATGTTTACGTTTATTGAAACTAGGGCAAAACTGGA 480  
Db 345288 GGAGAGAGAAAGCCAAAGGCTGAATGTTTATTGAAACTAGGGCAAAAGCTGGA 345229  
Qy 481 TACAAATGAAAGCAGCTCTTTTCGACGCTGAGCAGCTTTGCGGGAATGGAAGCACA 540  
Db 345228 TACAAATGAAAGCAGCTCTTTTCGACGCTGAGCAGCTTTGCGGGAATGGAAGCACA 345169  
Qy 541 CAGGACGGAAGCAGAGNAGACATGAGTACATAAACTGAAAGCCCTCAGGAGCAAAACA 600  
Db 345168 CAGGACGGAAGCAGAGNAGACATGAGTACATAAACTGAAAGCCCTCAGGAGCAAAACA 345109  
Qy 601 GTCAGCAAGGCGGTGTTCTCTCTACTCTCCATGTCATCTTCAACCCCTTCCTCAGAAG 660  
Db 345108 GTCAGCAAGGCGGTGTTCTCTCTACTCTCCATGTCATCTTCAACCCCTTCCTCAGAAG 345049  
Qy 661 CCCCCTTACTCTTTTCAATGACTGCAAGTGAATATTGGCTTGAAACCTTTTCCCTTCATTA 720  
Db 345048 CCCCCTTACTCTTTTCAATGACTGCAAGTGAATATTGGCTTGAAACCTTTTCCCTTCATTA 344989  
Qy 721 ATAACGTTTTGCAATTCATCATTTGCTGCCTGTCTCGTGGAG 761  
Db 344988 ATAACGTTTTGCAATTCATCATTTGCTGCCTGTCTCGTGGAG 344948

RESULT 2  
US-09-925-300-424  
; Sequence 424, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben

Query Match 73.4%; Score 559.6; DB 17; Length 740;  
Best Local Similarity 91.0%; Pred. No. 1.4e-165;  
Matches 606; Conservative 0; Mismatches 59; Indels 1;

541 CAGGACGGAGACGAGAGACATGAGTGACATATAAACTGGAAAGCCTCAGGACAAACA 600

Db 611 CAGGACGAAGCAGAGAAGATATGATTGACATAAACTGGAAGCCCTCAGGAGCAACA 670  
Qy 601 GTCAGCGAAGGGGTGTTCTCTGCTACTCTCCGATGTCATCTTCAACCCCTTCCTCAGAAG 660  
Db 671 GTCAGTGAAGAGGGCTGTTCTGCTGAATGTCTCCCTAGTCATCTTCAA-CCTTCTTCAGAAG 729  
Qy 661 CCCCCT 666  
Db 730 CTCAC 735

## RESULT 4

US-10-094-749-430

; Sequence 430, Application US/10094749

; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, KYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOTUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 430

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-094-749-430

Query Match 52.6%; Score 401; DB 17; Length 2456;

Best Local Similarity 100.0%; Pred. No. 3.5e-115;

Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 361 ATCCGCTAGTAGGAATAGAACAGATCTTCTGCACAGAGGCAAGTGTCAAGTTGAGGAG 420  
Db 202 ATCCGCTAGTAGGAATAGAACAGATCTTCTGCACAGAGGCAAGTGTCAAGTTGAGGAG 261  
Qy 421 GGAGAGGAAGCAAGGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAACTGGA 480  
Db 262 GGAGAGGAAGCAAGGGCTGAATGTTACGTTTATTGAACTAGGGCAAAACTGGA 321  
Qy 481 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGAGCTTTGCCGGGAATGGAAAGCACA 540  
Db 322 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGAGCTTTGCCGGGAATGGAAAGCACA 381  
Qy 541 CAGACGGAAGCAGAGAGACATGAGTGACATAAACTGGAAGCCCTCAGGAGCAAAACA 600  
Db 382 CAGACGGAAGCAGAGAGACATGAGTGACATAAACTGGAAGCCCTCAGGAGCAAAACA 441  
Qy 601 GTCAGCGAAGGGGTGTTCTCTGCTACTCTCCCATGTCTCAACCCCTTCCTCAGAAG 660  
Db 442 GTCAGCGAAGGGGTGTTCTCTGCTACTCTCCCATGTCTCAACCCCTTCCTCAGAAG 501

Qy 661 CCCCCTACTCTTCTTCAATTGACTGCAGTGTGAATATTGGCTTGAACTTTCCCTTCATT 720  
Db 502 CCCCCTTACTCTTCTTCAATTGACTGCAGTGTGAATATTGGCTTGAACTTTCCCTTCATT 561  
Qy 721 ATAAAGTTTTGCAATTTCATCATTTGCTGCCTGTCTCGTGGAG 761  
Db 562 ATAAAGTTTTGCAATTTCATCATTTGCTGCCTGTCTCGTGGAG 602

## RESULT 5

US-10-172-118-1656

; Sequence 1656, Application US/10172118

; Publication No. US20030224374A1

; GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue

; APPLICANT: He, Yudong

; APPLICANT: Linsley, Peter

; APPLICANT: Mao, Mao

; APPLICANT: Roberts, Chris

; APPLICANT: Van 't Veer, Laura

; APPLICANT: Van de Vijver, Marc

; APPLICANT: Bernards, Rene

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

; FILE REFERENCE: 9301-175-999

; CURRENT APPLICATION NUMBER: US/10/172,118

; CURRENT FILING DATE: 2002-06-14

; PRIOR APPLICATION NUMBER: 60/380,770

; PRIOR FILING DATE: 2002-05-14

; NUMBER OF SEQ ID NOS: 2699

; SEQ ID NO 1656

; LENGTH: 1266

; TYPE: DNA

; ORGANISM: Homo sapiens

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: NM 016577

; DATABASE ENTRY DATE: 2001-06-18

US-10-172-118-1656

Query Match 50.3%; Score 383.2; DB 17; Length 1266;

Best Local Similarity 75.6%; Pred. No. 1e-109;

Matches 475; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 1 ATGTCGCGGGCGGAGACTTCCGGAATCCGCTCAGGAAATTCAGCTGGTGTTCTCGGG 60  
Db 338 ATGTCGCGAGGGGAGATTTTGGGATCCACTGAGAAATTCAGTTGGTGTCTTGGGG 397  
Qy 61 GAGCAAGCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCACAAC 120  
Db 398 GAGCAGAGCGTCGGGAAGACGCTCTCTGATTACGAGGTTTCATGTACGACAGCTTCGACAAC 457  
Qy 121 ACCTATCAGCAATAATTGGCATTTGATCTTTTATCAAAAACTATGTACTTGGAGGATGGA 180  
Db 458 ACATACCAGGCAACCATTTGGGATTTGACTTCTTGTCAAAAACCATGTACTTGGAGGCCGC 517  
Qy 181 ACAATCGGGCTTCGGCTGTGGGATACGGCGGTCAGAAAGCTCTCCGTAGCCTCATTTCCC 240  
Db 518 ACGTTCGAGCTGCAGCTCTCGGACACAGCTGGTCAGAGAGGTTCCGCGAGCTGATCCCC 577  
Qy 241 AGGTACATCCGTGATTCTGCTGCAGCTGTAGTATTACGATATCAAAATGTTAACTCA 300  
Db 578 AGCTACATCCGGGACTCCACGGTGGCTGTGTGTGTACGACATCAAAATCTCAACTCC 637  
Qy 301 TTCCAGCAAACTCAAAAGTGGATTTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360  
Db 638 TTCCAAACAGACCTCTAAGTGGATTCGACGACGTCAGGACAGAGAGGGGAGTGTGTGTATC 697  
Qy 361 ATCACGCTAGTGTAGGAATAGACAGATCTTGTCTGACAGAGGCAAGTGTGTGTGTGTGT 420  
Db 698 ATCATGCTGTGTGGGCAACAAGACGGACCTGGCTGTATTAAGAGGAGATTAACATCGAGAG 757  
Qy 421 GGAGAGAGGAAGCAAGGGCTGAATGTTTACGTTTATTGAAACTAGGGGCAAAACTGGA 480

Db 758 GGGAGCAGCGGCCAAAGAACTGAGCGTCATGTTTCATTGAGACCAGTGGAGACTGGC 817  
QY 481 TACAATGTAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGAAATGGAAGCAC 540  
Db 818 TACACGTGAGCAGCTTTTTCGACGTGTGGCTCTACCCGGAATGAGATGTC 877  
QY 541 CAGGACGGAAGCAGAGACATGAGTGACATATAAACTCGAAAAGCCTCAGGAGCAACA 600  
Db 878 CAGGAGAAAAGCAAGAGGATGATCGACATCAAGCTGGACAAACCCCGAGGCCCCG 937  
QY 601 CTCAGCGAAGGGGTTGTTCTCTGCTACT 628  
Db 938 GCCAGCGAGGGCGGCTGCTCTGCTAAT 965

## RESULT 6

US-10-342-887-1656  
; Sequence 1656, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.

; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; PRIOR FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 2699

; SEQ ID NO 1656  
; LENGTH: 1266

; TYPE: DNA  
; ORGANISM: Homo sapiens

US-10-342-887-1656

Query Match 50.3%; Score 383.2; DB 17; Length 1266;  
Best Local Similarity 75.6%; Pred. No. 1e-109;  
Matches 475; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGTCGCGGGGAGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTGTCTCTGGGG 60  
Db 338 ATGTCGCGAGGGGAGATTTTGGGAATCCACTGAGAAATTCAGTTGGTGTCTTGGGG 397  
QY 61 GAGCAAGCGTTCGCAAGACATCTTTGATCACAGATTGAGTATGACAGTTTGCACAC 120  
Db 398 GAGCAGCGTCGGGAAGACGTCCTGATTACAGGTTTCATGTACGACAGCTTCGACAC 457  
QY 121 ACCTATCAGCAATATTTGGCATTTGACTTTTATCAAAACTATGTACTTGGAGGATGA 180  
Db 458 ACATACAGGCAACCATTTGGGATTTGACTTTTGTCAAAAACCATGTACTTGGAGGACCG 517  
QY 181 ACATATCAGCAATATTTGGCATTTGACTTTTATCAAAACTATGTACTTGGAGGATGA 180  
Db 458 ACATACAGGCAACCATTTGGGATTTGACTTTTGTCAAAAACCATGTACTTGGAGGACCG 517  
QY 181 ACATATCAGCAATATTTGGCATTTGACTTTTATCAAAACTATGTACTTGGAGGATGA 180  
Db 518 ACGTGGCACTGAGCTCTGGGACACAGCTGTGTCAGGAGGTTCCGCAAGCTGATCCCC 577  
QY 241 AGGTACATCCGTAATTTCTGCTGAGCTGTAGTATGATGTCAGAACAGAGAGAGTGTATTC 300  
Db 578 AGTACATCCGGAAGCTCTGGGACACAGCTGTGTCAGGAGGTTCCGCAAGCTGATCCCC 577  
QY 241 AGGTACATCCGTAATTTCTGCTGAGCTGTAGTATGATGTCAGAACAGAGAGAGTGTATTC 300  
Db 578 AGTACATCCGGAAGCTCTGAGCTGTGGGACACAGCTGTGTCAGGAGGTTCCGCAAGCTGATCCCC 577  
QY 301 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGAGAGTGTATTC 360  
Db 638 TTCCAAACAGACCTCTTAAGTGGATTCGACGAGCTGAGGACAGAGAGGGGAGTGTATTC 697

QY 361 ATCAGCTAGTAGAAATAGAACAGATCTTGCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420  
Db 698 ATCATGCTGTGGGCAACAAGACGGACCTGGCTGATAAGAGGCAGATAACCATCGAGGAG 757  
QY 421 GGAGAGAGAAAGCCAAAGGGCTGAATGTTAGCTTTTATTGAAACTTAGGGCAAAAATCGGA 480  
Db 758 GGGGAGCAGCGCGCCAAAGAACTGAGCGTCATGTTTCATTGAGACAGTTCGGAAGACTGGC 817  
QY 481 TACAATGTAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGAAATGGAAGCAC 540  
Db 818 TACAACGTGAAGCAGCTTTTTCGACGTGTGGCTCGGCTCTACCCGGAATGAGATGTC 877  
QY 541 CAGGACGGAAGCAGAGACATGAGTGACATATAAACTGGAAAAGCCTCAGGAGCAACA 600  
Db 878 CAGGAGAAAAGCAAGAGGATGATCGACATCAAGCTGGACAAACCCCGAGGCCCCG 937  
QY 601 CTCAGCGAAGGGGTTGTTCTCTGCTACT 628  
Db 938 GCCAGCGAGGGCGGCTGCTCTGCTAAT 965

## RESULT 7

US-10-848-755A-181

; Sequence 181, Application US/10848755A  
; Publication No. US20050054826A1  
; GENERAL INFORMATION:

; APPLICANT: Mao, Mao

; TITLE OF INVENTION: HUMAN DIAPHANOUS-3 GENE AND METHODS OF USE THEREFOR

; FILE REFERENCE: 9301-196-999

; CURRENT APPLICATION NUMBER: US/10/848,755A

; CURRENT FILING DATE: 2004-05-18

; PRIOR APPLICATION NUMBER: 60/471,842

; PRIOR FILING DATE: 2003-05-11

; NUMBER OF SEQ ID NOS: 275

; SOFTWARE: Patencin version 3.2 CAM: 301891-999188

; SEQ ID NO 181

; LENGTH: 1266

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-848-755A-181

Query Match 50.3%; Score 383.2; DB 19; Length 1266;  
Best Local Similarity 75.6%; Pred. No. 1e-109;  
Matches 475; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1 ATGTCGCGGGGAGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTGTCTCTGGGG 60  
Db 338 ATGTCGCGAGGGGAGATTTTGGGAATCCACTGAGAAATTCAGTTGGTGTCTTGGGG 397  
QY 61 GAGCAAGCGTTCGCAAGACATCTTTGATCACAGATTGAGTATGACAGTTTTCACAC 120  
Db 398 GAGCAGCGTCGGGAAGACGTCCTGATTACGAGTTCATGTACGACAGCTTCGACAC 457  
QY 121 ACCTATCAGCAATATTTGGCATTTGACTTTTATCAAAACTATGTACTTGGAGGATGA 180  
Db 458 ACATACAGGCAACCATTTGGGATTTGACTTTTGTCAAAAACCATGTACTTGGAGGACCG 517  
QY 181 ACATATCAGCAATATTTGGCATTTGACTTTTATCAAAACTATGTACTTGGAGGATGA 180  
Db 518 ACGTGGCACTGAGCTCTGGGACACAGCTGTGTCAGGAGGTTCCGCAAGCTGATCCCC 577  
QY 241 AGGTACATCCGTAATTTCTGCTGAGCTGTAGTATGATGTCAGAACAGAGAGTGTATTC 300  
Db 578 AGTACATCCGGAAGCTCTGAGCTGTGGGACACAGCTGTGTCAGGAGGTTCCGCAAGCTGATCCCC 577  
QY 301 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGAGAGTGTATTC 360  
Db 638 TTCCAAACAGACCTCTTAAGTGGATTCGACGAGCTGAGGACAGAGAGGGGAGTGTATTC 697  
QY 361 ATCAGCTAGTAGAAATAGAACAGATCTTGCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420  
Db 698 ATCATGCTGGTGGGCAACAAGACCGACCTGGCTGATAGAGGACAGATAACCATCGAGGAG 757

Qy	421	GGAGAGAGGAAAGCCAAAGGGCTGAATGTTATGTTTATTGAAACTAGGGCAAAACCTGGA	480
Db	758	GGGCGAGCAGCGGCCCAAAGAACTGAGCGTCATGTTTCATTGAGACAGTGCAGAACTGGC	817
Qy	481	TACAAATGTAAGCAGACTCTTTTCGACGTGTAGCAGCAGCTTTGCCGGAAATGGAAGCACA	540
Db	818	TACAACTGTAAGCAGACTTTTTCGACGTGTGGCGTCTACCCGGAAATGGAATGTC	877
Qy	541	CAGGACCGAAGCAGAGAAGACATGAGTGACATAAACTGGAAGAGCCTCAGGAGCAACA	600
Db	878	CAGGAGAAAGCAAGAGGGATGATGCACATCAAGCTGGACAACCCGAGGAGCCCCCG	937
Qy	601	GTACGCGAAGGGGGTTGTTCTCTGCTACT	628
Db	938	GCCAGCGAGGGCGGCTGCTCTCTGCTAAT	965

RESULT 8  
US-09-960-352-7355  
; Sequence 7355, Application US/09960352  
; Patent NO. US2002037139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Weasley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathalagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 7355  
; LENGTH: 443  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 32-LIB188-014-Q1-E1-H11  
US-09-960-352-7355

Query Match 50.1%; Score 382; DB 9; Length 443;  
Best Local Similarity 93.0%; Pred. NO. 1.3e-109;  
Matches 400; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy	166	TACTTGGAGGATGGAAACAATCGGGCTTCGGCTGTGGGATACGGCGGGTCAGGAAACGTCCTC	225
Db	14	TACTTGGAGATCGAAACAATCAGGCTGCAGCTGTGGGATACTGCGGGTCAGGAAACGTTTC	73
Qy	226	CGTAGCCTCATTTCCCAAGGTACATCCGTGATTTCTGCTGCAGCTGTAGTAGTTTACGATATC	285
Db	74	CGTAGCCTCATTTCCCAAGTTACATCCGTGATTTCTGCTGCAGCTGTAGTAGTTTATGATATC	133
Qy	286	ACAAATGTTTAACTCATTTCCAGCAAACTACAAGTGGATTGATGATGTCAGAAACAGAAAGA	345
Db	134	ACAAATGTTTAACTCCTTCAGCAAACTACAANAATGGATTGATGATGTCAGAAACAGAAAGA	193
Qy	346	GGAAAGTGATGTTTATCATCACGCTAGTAGGAAATAGAAACAGATCTTGCTCAAGAGGCAAA	405
Db	194	GGAAAGTGATGTTTATCATCATGCTAGTAGGAAATAAACAGATCTTGCTGCAGAGAGCAAA	253
Qy	406	GTGTCAATGTTGAGGAGGAGAGAGAAAGCCAAAGGGCTGAATGTTACGTTTATTTGAAATC	465
Db	254	GTGTCAATCGAGGAAGGAGAGAGAAAGCCAAAGAGCTGAATGTTATGTTTATTTGAAATC	313
Qy	466	AGGGCAAAAACCTGGATACAATGTAAGCAGCTCTTTTCGAGCTGTAGCAGCAGCTTTGCCG	525
Db	314	AGTGCAAAAAGCAGGATACAATGTAAGCAGCTCTTTCCGAGCTGTTCGAGCAGCTCTACCT	373
Qy	526	GGAAATGGAAGCACACAGCGGAAGCAGAGAAACATGATGATCAATAAACTGGAAAAG	585
Db	374	GGAAATGGAAGCACACAGGACAGAGCAAGACATGATTGACATATAAACTGGHAAA	433
Qy	586	CCTCAGGAGC	595

Db 434 CCTCAAGAGC 443

```
RESULT 9
US-10-017-161-1727
; Sequence 1727, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1727
; LENGTH: 3826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(3826)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(334)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2878)..(2995)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3081)..(3626)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3370)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3568)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3776)..(3826)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1727
```

Query Match	46.2%	Score 352.2	DB 15	Length 3826
Best local Similarity	93.1%	Pred. No. 1.2e-99		
Matches 443	Conservative 0	Mismatches 25	Indels 8	Gaps 7
Qy	1	ATGTCGCGGGCGGAGACTTTCGGGAATCCGCTCAGGAAATTCAGCTGTGTGTTCTCTCGGG	60	
Db	3302	ATGTCCCGGGC-GAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGTGTGTTCTCTCGGG	3360	
Qy	61	GAGCAAGCGTTGCAAGACATCTTTGATCACACAGATTTCAGGTATGACAGTTTTCACAAC	120	
Db	3361	GAGCAAA-CTNTGCAAGACATCTTTGATCACACAGATTTCAGGTATGACAGTTTTCACAAC	3419	
Qy	121	ACCTATCAGGCAATAATTTGGCAATTGACTTTTATCAAAAACTATGTACTTTGGAGGATGGA	180	
Db	3420	ACCTATCAGGCAATAATTTGGCAATTCCTTTTATCAAAAACTATGTACTTTGGAGGATGGA	3479	
Qy	181	ACAATCGGGCTTCGGCTGTGGGATACGGCGGTCAGGAACGTCTCGTAGCCCTCATTC	240	
Db	3480	ACAATCGGGCTTCGGCTGTGGGATACGGCGGTCAGGAACGTCTCGTAGCCCTCATTC	3539	
Qy	241	AGGTACATCCGTGATTCCTGCTGCAGCTGTAGTAGTTTACGATATCAAAATGTTAACTCA	300	
Db	3540	AGGTACATCCGTGATTCCTGCTGCAGCTGAGTAGTTTACGATATCAAAATGTTAACTCA	3599	



QY 301 TTCCAGCAACTA-CAAAGTGATTGA-TGATGTCAGAA-CAGAAAGAGGAAGTGATGTT 357  
Db 3600 TTCCAGCAGACTACCAAGTGATTGACTGATGTCAGAACAGAGGAGGAGTGATGTT 3659  
QY 358 ATCATCAGCTAGTAGGAATAAGAACAGATCTTGCTGCAAGAGGAGGAGTGATGTTGAG 417  
Db 3660 ATCATCAGCTAGGCAA--TAAACAGATCTTGCTGCAAGAGGAGGAGTGATGTTGAG 3717  
QY 418 GAGGAGAGAGGAGAAAGCCA-AAGGCTGAATGTTAGCTTTATTGAAACTAGGGGCAA 473  
Db 3718 GCCGGAGAGCAGAAAGCCACAGGAGAGACATGTTACCTTTATTGAAACTAGGGGCAA 3774

## RESULT 10

US-10-292-798-1383  
; Sequence 1383, Application US/10292798  
; Publication No. US2003023583A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1383  
; LENGTH: 3826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; LOCATION: source  
; FEATURE:  
; LOCATION: (1)..(3826)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (201)..(334)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2878)..(2995)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3081)..(3626)  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (3370)..(3370)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (3568)..(3568)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (3776)..(3826)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
US-10-292-798-1383

Query Match 46.2%; Score 352.2; DB 17; Length 3826;

Best Local Similarity 93.1%; Pred. No. 1.2e-99;  
Matches 444; Conservative 0; Mismatches 25; Indels 8; Gaps 7;

QY 1 ATGTCGGGGGGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60  
Db 3302 ATGTCGGGGGGG-GAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 3360  
QY 61 GAGCAAGCGTTTGCAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTTCACAA 120

Db 3361 GAGCAAA-CTNTGCAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTTCACAA 3419  
QY 121 ACCTATCAGGCAATAAATTTGGCATTGACATTTTATCAAAACTATCTACTTTGGAGATGGA 180  
Db 3420 ACCTATCAGGCAATAAATTTGGCATTGACATTTTATCAAAACTATCTACTTTGGAGATGGA 3479  
QY 181 ACAATCGGGCTTCGGCTGTGGGATACGGGGGTGAGGAACGCTCTCCGTAGCCTCAATCCC 240  
Db 3480 ACAATCGGGCTTCGGCTGTGGGATACGGGGGTGAGGAACGCTCTCCGTAGCCTCAATCCC 3539  
QY 241 AGGTACATCCGTGATCTTGCTGAGCTGTAGTGTAGTGTACGATATCAGAAATGTTAACTCA 300  
Db 3540 AGGTACATCCGTGATCTTGCTGAGCTGTAGTGTAGTGTACGATATCAGAAATGTTAACTCA 3599  
QY 301 TTCCAGCAACTA-CAAAGTGATTGA-TGATGTCAGAA-CAGAAAGAGGAGTGATGTT 357  
Db 3600 TTCCAGCAGACTACCAAGTGATTGACTGATGTCAGAACAGAGGAGGAGTGATGTT 3659  
QY 358 ATCATCAGCTAGTAGGAATAAGAACAGATCTTGCTGCAAGAGGAGGAGGAGGAGGAGG 417  
Db 3660 ATCATCAGCTAGTAGGCAA--TAAACAGATCTTGCTGCAAGAGGAGGAGGAGGAGG 3717  
QY 418 GAGGAGAGAGGAGAAAGCCA-AAGGCTGAATGTTAGCTTTATTGAAACTAGGGGCAA 473  
Db 3718 GCCGGAGAGCAGAAAGCCACAGGAGAGACATGTTACCTTTATTGAAACTAGGGGCAA 3774

## RESULT 11

US-09-918-995-5540  
; Sequence 5540, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5540  
; LENGTH: 424  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(424)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-5540

Query Match 44.9%; Score 341.8; DB 10; Length 424;

Best Local Similarity 89.8%; Pred. No. 6.3e-97;  
Matches 380; Conservative 0; Mismatches 31; Indels 12; Gaps 1;

QY 334 AGAACAGAAAGAGGAAGTGATGTTATCATCAGCTAGTAGGAAATAGAACAGATCTTGC 393  
Db 1 AGAACAGAAAGAGGAAGNGATGNTATCANGCTAGTTTTAAATPAAACAGATCTTGC 60  
QY 394 GACAAGAGGCAAGTGTCAGTTTCAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 453  
Db 61 GACAAGAGGCAAGTGTCAGTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
QY 454 TTTATTGAACTAGGGGCAAACTGGATACAAATGTAAGAGAGCTCTTTCGAGCTAGCA 513  
Db 121 TTTATTGAACTAGTGCAAAAGCTGGATACAAATGTAAGAGAGCTCTTTCGAGCTAGCA 180  
QY 514 GCAGCTTTGCCGGGATGGAAGCAGCAGGCGGAGCAGAGAGAGATGAGTGCACATA 573  
Db 181 GCAGCTTTGCCGGGATGGAAGCAGCAGGCGGAGCAGAGAGAGATGAGTGCACATA 240

Qy 574 AAATGGAAAGCCTCAGGCGAACAACAGTACAGGAAGGGGTGTTCTCTGCTACTCTCCC 633  
Db |||||  
Qy 241 AAATGGAAAGCCTCAGGCGAACAACAGTACAGGAAGGGGTGTTCTCTGCTACTCTCCC 300  
Db |||||  
Qy 634 ATGTCATCTTCAACCCCT-----TCTCAGAAAGCCCTTACTCTCTTTCATTGAC 681  
Db |||||  
Qy 301 ATGTCATCTTCAACCTTCTTCAAGAGCTCACTGCTTTGGCCCTTACTCTTTCATTGAC 360  
Db |||||  
Qy 682 TGCAGTGTGAATATTTGGCTTGAACCTTTTCCCTTCATTAATAAGCTTTTGGCAATTCATCA 741  
Db |||||  
Qy 361 TGCAGTGTGAATATTTGGCTTGAACCTTTTCCCTTCAGTAATAAGCTATTGCAATTCATCA 420  
Db |||||  
Qy 742 TTG 744  
Db |||||  
Qy 421 TTG 423  
Db |||||  
RESULT 12  
US-10-108-605-44  
; Sequence 44, Application US/10108605  
; Publication No. US20020160934A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadus, Julie  
; APPLICANT: Stam, Lynn  
; APPLICANT: Bachmann, Jane  
; APPLICANT: Kamdar, Kim  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
; FILE REFERENCE: 31133B  
; CURRENT APPLICATION NUMBER: US/10/108,605  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 09/761,142  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/176,418  
; PRIOR FILING DATE: 2000-01-14  
; NUMBER OF SEQ ID NOS: 361  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 44  
; LENGTH: 1157  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-10-108-605-44

Query Match 39.1%; Score 298; DB 13; Length 1157;  
Best Local Similarity 68.8%; Pred. No. 7.7e-83;  
Matches 425; Conservative 0; Mismatches 190; Indels 3; Gaps 1;  
Qy 12 CGGAGACTTCGGGAATCCGCTGAGGAATTCAGTGTGTCTCTGGGGAGCAAGCGT 71  
Db |||||  
Qy 300 CGGAGATTTGGCAATCCGCTGCGGAAGTTCAAGCTGTCTCTCTCGCGAGCAGATGT 359  
Db |||||  
Qy 72 TGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCACACACCTATCAGGC 131  
Db |||||  
Qy 360 GGGCAAGACCTCGCTGATTACAGATTTCATGTACGACAGCTTCGACAAACGTTACCAAGC 419  
Db |||||  
Qy 132 AATTAATGGCAATTCATTTTATCAAAACATATGTACTTGGAGATGGAAACATCGGGCT 191  
Db |||||  
Qy 420 GACGATCGGGAATGATTTCTTATCGAAGACATGTACTTGGAGATCGCACTGTGCGCCT 479  
Db |||||  
Qy 192 TCGGCTGTGGGATACGGCGGTGAGGAACGCTCTCGTAGCTCATTTCCAGGTATACATCCG 251  
Db |||||  
Qy 480 CGAGCTGTGGGATACGGCGGAGCAGGAGCGATTCGCTCGCTGATACCTCGTACATACG 539  
Db |||||  
Qy 252 TGATTTCTGCTCAGCTGTAGTTTACGATATCAAAATGTTAACTCATTTCCAGCAAC 311  
Db |||||  
Qy 540 CGATCCACGGTGGCAGTGGTGGTTACGATATCAACCAACCACTCGTTCCACCAAGC 599  
Db |||||  
Qy 312 TACAAGTGGATTCATGTATGTCAGAACAGAGGAGTATGTTATCATCATCGCTAGT 371  
Db |||||  
Qy 600 CTCAAGTGGATTCATGTATGTCAGAACAGAGGAGTATGTTATCATCATCGCTAGT 659  
Db |||||  
Qy 372 AGGAATATGAACAGATCTTGTGTAAGAGGCAAGTGTCACTTGGAGGGGAGAGGAA 431  
Db |||||

Db 560 GGGCAACAAGACGATCTCTCCGACAAGCGTCAGTGTCCACCGAGGAGGGTGAGCGCA 719  
Qy 432 AGCCAAAGGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAAACTGGATACATGTAAA 491  
Db |||||  
Qy 720 GGGCAAGGAGCTTAACGTTGATGTTTCATCGAGACAAGCGCAAGCGCGGTACATGTGAA 779  
Db |||||  
Qy 492 GCAGCTCTTTTCGAGCTGTAGCAGCAGCTTTGCCGGGAATGGAAGACACACAGGACGGAAG 551  
Db |||||  
Qy 780 GCAATTTTCCGACGGGTGGCGGCACCTGCGGCATGGAATTCACCGAGAAACAAGCC 839  
Db |||||  
Qy 552 CAGAGAAACATGATGACATAAAACT---GGAAAAGCCTCAGGAGCAAAACAGTCAAGCA 608  
Db |||||  
Qy 840 CTCCGAGGACATGCAGGAGGTGTGCTAAAGGACTCACCAACGAGACAAAGGATCCGA 899  
Db |||||  
Qy 609 AGGGGTTGTTCTGTGCTA 626  
Db |||||  
Qy 900 GGGCGGCTGGCGCTGCTA 917  
Db |||||

RESULT 13  
US-10-437-963-63907/c  
; Sequence 63907, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 63907  
; LENGTH: 1175  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MPT4530\_65103C.1  
US-10-437-963-63907

Query Match 35.0%; Score 266.8; DB 18; Length 1175;  
Best Local Similarity 67.0%; Pred. No. 5.5e-73;  
Matches 379; Conservative 0; Mismatches 187; Indels 0; Gaps 0;  
Qy 16 GACTTGGGAATCCGCTGAGGAATTCAGCTGGTGTTCCTGGGGAGCAAGCGTTGCA 75  
Db 1080 GCCCCCGTGTGCGCGCTCGCAAGTACAGCTGGTGTTCCTCGGCGACCAGTCCGTCGGC 1021  
Qy 76 AAGACATCTTTGATCACCAGATTACAGGTATGACAGTTTTCACACACCTATCAGGCAATA 135  
Db 1020 AAGACAGCATATCACCCTTCATGTACGACAGTTTCACACACCTACAGGCTACA 961  
Qy 136 ATTGGCATTCGACTTTTTATCAAAAACATATGCTTGGAGATGGAAACATCGGCTTCGG 195  
Db 960 ATCGGTATTGACTTCTGTCAAGACAATGTACCTTGAAGATAGGACTGTGAGACTGCAA 901  
Qy 196 CTGTGGGATACGGGGGTGAGGAACGCTTCGTTAGCTCATTTCCAGGTACATCCGCTGAT 255  
Db 900 CTTTGGGATACAGCTGGTCAAGAACGATTTCAGGAGTTTGAATCCAGCTATATCAGAGAC 841  
Qy 256 TCTGCTCAGCTGTAGTGTGTTACGATATCAGAAATGTTAACTCATTTCCAGCAAACTACA 315  
Db 840 TCTTCCGCTGCTGTCATCGTATTGTTGTTGAGAGCAGACAGCTTTTCTTAACACTTCA 781  
Qy 316 AAGTGGATTGATGTGACAGAAAGAGGAGTGTATATCATCATCGCTAGTAGGA 375  
Db 780 AATGGATAGAGGAAGTTAGAACTGAGAGGGGCGAGTGTATGTTATCATTTGTGCTGTTGGG 721



Db       861 CTG 863  
|||

Search completed: April 25, 2005, 13:07:29  
Job time : 526 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2005, 06:53:31 ; Search time 3096 Seconds  
(without alignments)  
9368.531 Million cell updates/sec

Title: US-09-441-857-11  
Perfect score: 762  
Sequence: 1 atgtccggcgaggagactt.....tgctgcctgtctcgtggaga 762

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_hic:\*
- 4: gb\_est3:\*
- 5: gb\_est4:\*
- 6: gb\_est5:\*
- 7: gb\_est6:\*
- 8: gb\_gsa1:\*
- 9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	761	99.9	3084	3 HSM801695	AL136727 Homo sapi
2	596.6	78.3	1928	3 AK084131	AK084131 Mus muscu
3	596.6	78.3	1955	3 AK051246	AK051246 Mus muscu
4	596.6	78.3	3404	3 AK083262	AK083262 Mus muscu
5	595	78.1	897	5 BQ422263	BQ422263 AGENCOURT
6	586	76.9	628	6 CB565915	CB565915 AGENCOURT
7	576.6	75.7	615	6 CB067426	CB067426 iq36f10.x
8	570.6	74.9	609	5 BQ632282	BQ632282 il2sa01.x
9	566	74.3	790	4 BI694279	BI694279 603347724
10	565.4	74.2	599	7 CV025778	CV025778 3550 Full
11	544	71.4	656	1 AI435940	AI435940 th80c04.x
12	528	69.3	744	6 CB320435	CB320435 AGENCOURT
13	525.2	68.9	762	2 BE965054	BE965054 601658839
14	522.6	68.6	632	2 BF431157	BF431157 7008c05.x
15	512	67.2	792	6 CA316338	CA316338 UI-M-FW0-
16	499	65.5	809	7 CN227000	CN227000 RJB008D12
17	492.4	64.6	637	4 BI602796	BI602796 603247154
18	490.2	64.3	554	1 AI660753	AI660753 we66h09.x
19	480.2	63.0	600	4 BG805486	BG805486 0693-37 M
20	470	61.7	871	5 BU381224	BU381224 603857076
21	469.2	61.6	880	4 BI602882	BI602882 603250358
22	469	61.5	769	6 CB988054	CB988054 AGENCOURT
23	467.4	61.3	648	7 CF172759	CF172759 B0912F09-
24	464.4	60.9	530	5 BQ086790	BQ086790 ih90b06.Y

25	464.4	60.9	534	5 BQ086839	BQ086839 ih90g05.Y
26	461.8	60.6	585	2 AW975757	AW975757 EST387866
27	461.6	60.6	518	4 BM833772	BM833772 K-EST0108
28	460.2	60.4	764	5 BU358285	BU358285 603476202
29	458.6	60.2	914	6 CA974052	CA974052 AGENCOURT
30	457.6	60.1	838	7 CK315856	CK315856 SB02028A1
31	457.4	60.0	572	7 CF198211	CF198211 maj45e04.
32	452.6	59.4	549	1 AL600766	AL600766 DKFZp313C
33	452	59.3	532	1 AA824586	AA824586 OC83C09.B
34	450.4	59.1	614	5 BX279398	BX279398 BX279398
35	449	58.9	563	7 CR556809	CR556809 DKFZp455E
36	447.6	58.9	804	7 CK302212	CK302212 SB02015A1
37	447.6	58.7	696	5 BX852687	BX852687 BX852687
38	446	58.5	883	4 BG108074	BG108074 602280008
39	445	58.4	655	8 AZ609020	AZ609020 IM0433M03
40	444.6	58.3	757	7 CK302742	CK302742 SB02016A2
41	444.4	58.3	878	6 CD361397	CD361397 AGENCOURT
42	441.2	57.9	858	7 CF547798	CF547798 AGENCOURT
43	439.4	57.7	759	5 BF709291	BF709291 BP709291
44	439.4	57.7	775	1 AI893257	AI893257 me39c01.Y
45	438.4	57.5	886	5 BU363505	BU363505 603785110

ALIGNMENTS

RESULT 1  
HSM801695 HSM801695 Homo sapiens mRNA; cDNA DKFZp566K144 (from clone DKFZp566K144).  
LOCUS AL136727  
DEFINITION AL136727.1 GI:12052972  
ACCESSION HTCC  
VERSION HTCC  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3084)  
AUTHORS Bloecker, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.  
CONSRTH The German cDNA Consortium  
TITLE Direct Submission  
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY  
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp566K144) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp566K144  
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.  
FEATURES  
source  
1. 3084  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp566K144"  
/db\_xref="taxon:9606"  
/clone="DKFZp566K144"  
/tissue type="kidney"  
/clone\_lib="566 (synonym: hfk22). Vector pAMP1; host XI-2blue; sites NotI + SalI"  
/dev stage="fetal"  
/note="RAB6C, member RAS oncogene family"  
1. 3084  
/gene="DKFZp566K144"  
456. 1220  
/gene="DKFZp566K144"  
/codon\_start=1  
/product="hypothetical protein"



```
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
Source
1..1928
/organism="Mus musculus"
/mol_type="mrna"
/strain="C57BL/6J"
/db_xref="FANTOM DB:DI130097N11"
/db_xref="taxon:10090"
/clone="DI130097N11"
/tissue_type="spinal ganglion"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
447..1073
/note="unnamed protein product; RAB6, member RAS oncogene family [MGD|MGI:894313, GB|NM_024287, evidence: BLASTN, 99%, match=1919] putative"
/codon_start=1
/protein_id="BAC39121.1"
/db_xref="GI:26350969"
/translation="MSAGDGFNPLRKLFLVFGSVGKTSLLITREMYDSFDNTYQA TIGIDFLSKTYLEBTRVLQWLDTAGQERFLPSYIRSTVAVVYDITNNSFQ QTKWIDVTRGSDVIIMLVGNKTDLDADKRVSEIEGERRAKELNWMFIETSAKAG YNVKQIFRVAALPGMESTQDRSDREDMIDIKLEKQPQPVNEGSC"
CDS
1..1928
/query Match 78.3%; Score 596.6; DB 3; Length 1928;
Best Local Similarity 86.9%; Pred. No. 1.3e-167;
Matches 672; Conservative 0; Mismatches 89; Indels 12; Gaps 1;
ORIGIN
1 ATGTCGGCGGGCGGAGACTTCGGGAATCGCTGAGGAATTCGAAGTGTGTCTCTGGGG 60
Db 447 ATGTCGGCGGGCGGAGACTTCGGGAATCGCTGAGGAATTCGAAGTGTGTCTCTGGGA 506
Qy 61 GAGCAAGCGTTCGCAAGACATCTTTGATCACAGATTGAGTATGACAGTTTTCGACAC 120
Db 507 GAGCAGAGCGTTGGAAAGACGCTTGTATGATCAGCCGATTCATGATGACAGTTTTCGACAC 566
Qy 121 ACCTATCAGGCATTAATGGCATTCGCTTTTATCAAAACATCTGCTACTGGAGATGGA 180
Db 567 ACCTATCAGGCATTAATGGCATTCGCTTTTATCAAAACATCTGCTACTGGAGATGGA 626
Qy 181 ACAATCGGGCTTCGGGTGGGATACGGCGGTTCAGGAACGCTCTCGTAGCCCTCATTC 240
Db 627 ACGTGGCATTCGAATATGGGACACAGCAGCTCAAGCGGTTCAGGAGCTTGATTCCT 686
Qy 241 AGGTATATCCGTGATTTCTCTGAGCTGTAGTGTATGATATACAAATGTTAACTCA 300
Db 687 AGGTATATCCGTGATTTCTCTGAGCTGTAGTGTATGATATACAAATGTTAACTCA 746
Qy 301 TTCCAGCAAACTACAAAGTGGATTCGATGTCAGCAAGAGAGGAGTGTATCTTATC 360
Db 747 TTCCAGCAAACTACAAAGTGGATTCGATGTCAGCAAGAGAGGAGTGTATCTATC 806
Qy 361 ATCAGCTAGTAGGAATAGAACAGATCTTGTGACAGAGCAAGTGTCTGAGGAG 420
Db 807 ATCATGCTAGTAGGAATAGAACAGATCTTGTGATAGAGCAAGTGTCTGAGGAG 866
Qy 421 GGAGAGAGAAAGCCAAAGGCTGAATGTATGTTTATTTGAACATAGGCAAAATCTGA 480
Db 867 GGAGAGAGAAAGCCAAAGGCTGAATGTATGTTTATTTGAACATAGGCAAAAGCAGA 926
Qy 481 TACAATGTAAAGCAGCTCTTTCGAGTGTAGCAGCAGCTTTCGCGGGAATGAAAGCACA 540
Db 927 TACAATGTAAAGCAGCTTTCGCGGCTGTGTGTCAGCTTTTACCTGGAATGAAAGCACA 986
Qy 541 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 987 CAGGACGGAAGCAGAGAACATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
Qy 601 GTCCAGGAAGGGGGTGTTCCTGCTACTCTCCCATGTCTTCAACCCCTCTCCAGAG 660
```

```
Db 1047 GTCAATGAAGCGCGCTGTCTCTGCTAATCCCTGGCATCTTTCAACCCCTCTCCAGAG 1106
Qy 661 C-----CCCTTACTCTTCAATGACGTCAGTGTGAATATTGGCTTGAACCTT 708
Db 1107 CTCACTGCTTTGGGCCCATCTCTTCAATGACGTCAGTGTGAATATTGGCTTGAACCTT 1166
Qy 709 TTCCCTTCAATTAAGCTTTTGAATTCATCATCTGCTGCTCTCTCTGAG 761
Db 1167 TTCCCTTCAATTAAGCTTTTGAATTCATCATCTGCTGCTCTCTCTGAG 1219

RESULT 3
AK051246
LOCUS
DEFINITION
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone:DI130023A03 product:RAB6, member RAS oncogene family, full insert sequence.
AK051246
VERSION
AK051246.1 GI:26341819
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
REFERENCE
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., Sasaki,N., Carninci,P., Suno,H., Akiyama,J., Nishi,K., Kusunagi,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
6 (bases 1 to 1955)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saio,R., Saichou,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akashira,S.,
```

Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
FEATURES  
Location/Qualifiers  
1..1955  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="PANTOM DB:D130023A03"  
/db\_xref="taxon:10090"  
/clone="D130023A03"  
/tissue\_type="spinal ganglion"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stages="12 days embryo"  
474..1100  
/note="unnamed protein product: RAB6, member RAS oncogene family [MGP|MGI:894313, GB|NM\_024287, evidence: BLASTN, 99%, match=1919]  
putative"  
/codon\_start=1  
/protein\_id="BAC34572.1"  
/db\_xref="GI:26341820"  
/translation="MSAGDSDGNLRFKFLVPLGEQSVGKTSLTRFMYSFDNTYQA  
TIDFIDFVTLWEDRVLQWMDTAGQERFSLIPSIYRSTDAVVYDITNNSFQ  
QTGWIDVTERGSDVIMLVGNKTLADLRQVSIIEGERKALNMVFIETSAKAG  
YNVQLFRVAAALPGMESTQDRSREDMIDIKLEKPOEPVNEGCSG"  
ORIGIN  
Query Match 78.3%; Score 596.6; DB 3; Length 1955;  
Best Local Similarity 86.9%; Pred. No. 1.3e-167;  
Matches 672; Conservative 0; Mismatches 89; Indels 12; Gaps 1;  
QY 1 ATGTCGGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTCTCTCTGGGG 60  
DB 474 ATGTCGGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTCTCTCTGGGA 533  
QY 61 GAGCAAGCGTTGCAAGACATCTTTGATGATACCAAGATTCAGGTATGACAGTTTTCAGAAC 120  
DB 534 GAGCAGACGTTGGAAGACGCTCTTGATCAACCGGATTCATGATGACAGTTTTCAGAAC 593  
QY 121 ACCTATCAGGCAATATTTGGCATTCACATTTTATCAAAAATGATGATCTGAGGATCGA 180  
DB 594 ACCTATCAGGCAACAAATTTGGCATTCACATTTTATCAAAAATGATGATCTGAGGATG 653  
QY 181 ACAATCGCGGCTTCGGCTGCGGATACGGCGGCTCAGGAACGCTCTCCGCTGAGCTCATTC 240  
DB 654 ACCGTGCGATTCGAATATGAGACACAGCAGGTCAAGACGGTTCAGGAGCTTGATTCCT 713  
QY 241 AGGTATACCGGTGATCTGCTGAGCTGTAGTGTTCAGTATACAAATGTTAACTCA 300  
DB 714 AGCTACATTCAGAGACTCCACTGTGGCAGTTGTTGTTATGATATACAAATGTTAACTCA 773  
QY 301 TTCCAGCAAACTACAAAGTGGATTCATGATCTCAGACAGAAAGGAGGTGATGTTATC 360  
DB 774 TTCCAGCAAACTACAAAGTGGATTCATGATCTCAGACAGAAAGGAGGTGATGTTATC 833  
QY 361 ATCAGCGTGTAGGAAATAGAACAGATCTTCTGTCAGAGGCGCAAGTGTCTAGTTGAGGAG 420

Db 834 ATCATGCTAGTAGGAAATAAAACAGATCTTCTGATGAAGAGCAAGTGTCTCAATTGAGGAG 893  
QY 421 GGAGAGAGGAAAGCAAGGGCTGAATGTTACGTTTATTTGAAACTAGGGCAAAACTGGA 480  
Db 894 GGAGAGAGGAAAGCAAGAGCTGAATGTTATGTTTATTTGAAACAGTGTGCAAAAGCAGGA 953  
QY 481 TACAATGTAAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTCCGCGGAATCGGAAGACACA 540  
Db 954 TACAATGTAAAGCAGCTTTTCCGCGCTGTGCTGCAGCTTTTACCTTGAATGGAAGACACA 1013  
QY 541 CAGGACGAGAGCAGAGAGACATGAGTGACATAAACTGGAAAAAGCCTCAGAGCAACAACA 600  
Db 1014 CAGGACGAGAGCAGAGAGACATGATGACATAAACTGGAAAAAGCCTCAGAGCAACAACA 1073  
QY 601 GTGACGAGAGGGGTTGTTCTCTGCTACTCTCTCCCATGTCATCTTCAACCCCTTCCTCAGAAG 660  
Db 1074 GTCAATGAAGGCGCTGTCTCTGCTTAATCCCTCGGCATCTTTTCAACCCCTCTCGAGAAG 1133  
QY 661 C-----CCCCCTTACTCTTTCATTCGACTGCGAGTGTGAATATTGGCTTCAACCTT 708  
Db 1134 CTCACTGCTTTGGCCCCCATCTCTTTCATGACTGCGAGTGTGAATATTGGCTTGAACCTT 1193  
QY 709 TTCCCTTTCATTAATACGTTTTCGCAATTCATCATTCGCTGCTGCTGCTGCTGAG 761  
Db 1194 TTCCCTTTCAGTAATACGTTTTCGCAATTCATCATTCGCTGCTGCTGCTGCTGAG 1246  
RESULT 4  
AK083262 3404 bp mRNA linear HTC 03-APR-2004  
LOCUS  
DEFINITION Mus musculus adult male hippocampus cDNA, RIKEN full-length enriched library, clone: C630031P05 product: RAB6, member RAS oncogene family, full insert sequence.  
ACCESSION AK083262  
VERSION AK083262.1 GI:26350388  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
93279253  
10349636  
REFERENCE 2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159  
REFERENCE 3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
REFERENCE 4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)



AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
	Nature 420, 563-573 (2002)	
JOURNAL	6 (bases 1 to 3404)	
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.	
	Direct Submission	
TITLE	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	
FEATURES	1..3404	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
source	/strain="C57BL/6J"	
	/db_xref="FANTOM DB:C630031P05"	
	/db_xref="taxon:10090"	
CDS	/clone="C630031P05"	
	/sex="male"	
	/tissue_type="hippocampus"	
polyA_signal	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
	/dev_stage="adult"	
	463..1089	
polyA_site	/note="unnamed protein product; RAB6, member RAS oncogene family (MGD MG1:894313, GB NM_024287, evidence: BLASTN, 99%, match=1919)	
	putative	
	/codon_start=1	
ORIGIN	/protein_id="BAC38834.1"	
	/db_xref="GI:26350389"	
	/translation="MSAGDFGNPLRKLFLGDSQVSKTSLITRFMYDFDNTYQA TIGDLSKTYMLEDTRLVRLTAGQERFSLIPSYIRDSITVAVVYDITNVNSFQ QTKWIDVTRGSDVIIMLVGNKTDLADKQVSI EBERKAKELNWFETSAKAG YNVKQLFRVAALPGMESTQDRSDRMIDIKLXPQPYNEGSCS"	
Query Match	3379..3384	
	/note="putative"	
	3404	
Best Local Similarity	/note="putative"	
	87.3%; Score 596.6; DB 3; Length 3404;	
	Pred. No. 1.6e-167;	
Matches 672; Conservative 0; Mismatches 89; Indels 12; Gaps 1;		
1 ATGTCCGGCGGCGGACCTTCGGGAATCCGCTGAGGAAATTCAGCTGTGTCTCTGGG 60		
463 ATGTCCGGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGTGTCTCTGGG 522		
61 GAGCAAGCGTTGCAAGACATCTTTTGATCACCAGATTTCAGGTATGACAGTTTGGACAAC 120		
523 GAGCAGCGTTGGAAAGACGCTTGTATCACCAGATTTCATGTATGACAGTTTTCACAAC 582		
121 ACCTATCAGCAATTAATTTGGCATTGACCTTTTATCAAAAATATGTACTTGGAGATGA 180		
583 ACCTATCAGGCAACAATTTGGCATTGACCTTTATCAAAAACAATGTACTTGGAGATGA 642		
181 ACAATCGGGCTTCGGCTCTGGGATACGGCGGTTCAGGAACGCTCTCCGTAGCCTCAATCCC 240		
643 ACCGTGCGATTGCAATTATGGGACACAGAGTCAAGAGCGGTCAGGAGCTTGATCTCT 702		
241 AGGTATACCTCGTGAATCTTGCTGACGCTTAGTGTATAGTATACAAAATGTTAACTCA 300		
703 AGCTATCTCGAGACTCCACTGTGGCAGTGTGTGTATGATATCACAATGTTAACTCA 762		
301 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGAGGAGTGTATGTC 360		
763 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAGAGGAGTGTATGTCATC 822		
361 ATCAGCTAGTAGGAAATAGAACAGATCTTCTGCAAGAGGCAAGTGTCTAGTTGAGGAG 420		
823 ATCATGCTAGTAGGAAATAGAACAGATCTTCTGATAGAGGCAAGTGTCTCAATTGAGGAG 882		
421 GGAGAGAGAAAGCCAAAGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAACTGGA 480		
883 GGAGAGAGAAAGCCAAAGGCTGAATGTTATGTTATTGAAACCCAGTGCAGGAGCAGGA 942		
481 TACATGTAAAGCAGCTCTTTTCGACGCTGACAGCAGCTTTCCGGGATGGAAGACACA 540		
943 TACATGTAAAGCAGCTCTTTTCGACGCTGTTCCGGGCTTTACCTGGGAATGGAAGACACA 1002		
541 CAGGACGGAAGCAGAGAGACATGATGACATAAACTGGAAGAGCCCTCAGGAGCAAAACA 600		
1003 CAGGACGGAAGCAGAGAGACATGATGACATAAACTGGAAGAGCCCTCAGGAGCAAAACA 1062		
601 GTCAGGAGAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCCCTCTCTCAGAG 660		
1063 GTCAATGAAGCGGCTGTCTCTGCTTAATCCCGCTGTCATCTTCAACCCCTCTCTCAGAG 1122		
661 C-----CCCTTACTCTTTCATTGACGCTGTCAGTGAATATTGGCTTGAACCTT 708		
1123 CTCAGTCTTGGGCCCCATCTTTCAATGACGCTGTCAGTGAATATTGGCTTGAACCTT 1182		
709 TTCCCTTCATTAATAACGTTTTTGCAATTCATTCCTGCTGCTGCTGCTGAGGAG 761		
1193 TTCCCTTCAGTAATAACGTTATGCAATTTCATTCATTCCTGCTGCTGCTGAGGAG 1235		
RESULT 5	BO422263	
	BO422263	
	BO422263	
*ACCESSION	BO422263.1	
	GI:21117590	
	EST.	
KEYWORDS	Homo sapiens (human)	
	Homo sapiens	
	Homo sapiens	
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	NIH-MGC http://mgc.nci.nih.gov/	
	National Institutes of Health, Mammalian Gene Collection (MGC)	
REFERENCE	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-remail.nih.gov	
AUTHORS	Tissue Procurement: ATCC	
	cDNA Library Preparation: Life Technologies, Inc.	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
TITLE	DNA Sequencing by: Agencourt Bioscience Corporation	
	Clone distribution: MGC clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LLNL at:	
JOURNAL	http://image.llnl.gov	
COMMENT		



QY 301 TTCACGCAAACTACAAAGTGGATTGATGTCAGAACAGAAAGAGGAAGTGTATTC 360  
 Db 301 TTCACGCAAACTACAAAGTGGATTGATGTCAGAACAGAAAGAGGAAGTGTATTC 360  
 QY 361 ATCACGCTAGTAGGAATAGACAGATCTTGCTGACAAAGAGCAAGTGTCACTGAGGAG 420  
 Db 361 ATCATGCTAGTAGGAATAGACAGATCTTGCTGACAAAGAGCAAGTGTCAATGAGGAG 420  
 QY 421 GGAGAGAGGAAGCCAAAGGCTGAAATGTTAGCTTTATTGAAACTAGGGCAAAACTGGA 480  
 Db 421 GGAGAGAGGAAGCCAAAGAGCTGAATGTTATTGAAACTAGTCAAAAGCTGA 480  
 QY 481 TACAATGTAAGACAGCTCTTTCGACCTGTAGACAGCAGCTTTGCCGGGAATGGAAGACACA 540  
 Db 481 TACAATGTAAGACAGCTCTTTCGACCTGTAGACAGCAGCTTTGCCGGGAATGGAAGACACA 540  
 QY 541 CAGGACGGAAGCAGAGACACATGAGTGACATAAACTGGAAGAGCCTCAGGAGCAACACA 600  
 Db 541 CAGGACGGAAGCAGAGACATGATGTTGACATAAACTGGAAGAGCCTCAGGAGCAACCA 600  
 QY 601 GTCAGCGAAGGGGGTGTTCCTGCTA 626  
 Db 601 GTCAGTGAAGGAGGCTGTTCCTGCTA 626

RESULT 7  
 CB067426  
 LOCUS  
 DEFINITION  
 iQ35f10.x1 HR05 islet Homo sapiens cDNA clone IMAGE: 3' similar to  
 SW:RAB6\_HUMAN P20340 RAS-RELATED PROTEIN RAB-6. [1] ; mRNA  
 sequence.  
 CB067426  
 CB067426.1 GI:27811946

EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 615)  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,  
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,  
 Cardenas,M., Gibbons,M., McCann,R., Cole,R., Teagareishvili,R.,  
 Williams,T., Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@im.wustl.edu)  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 444.  
 Location/Qualifiers

FEATURES  
 Source  
 1..615  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:"  
 /tissue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /clone\_lib="HR05 islet"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
 NotI; Site\_2: XhoI; cDNA made by oligo-dT priming."

Size-selected on agarose gel. Average insert size ~1kb. 5'  
 xhoI site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,  
 Metabolism Div. (Alan Permutt Lab), Washington University  
 School of Medicine, Box 8127, 660 South Euclid Ave., St.  
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
 314-362-1916, Fax: 314-747-2692."

## ORIGIN

Query Match 75.7%; Score 576.6; DB 6; Length 615;  
 Best Local Similarity 96.1%; Pred No. 9.7e-162;  
 Matches 591; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 10 GSCGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTGTTCCTCGGGGAGCAAGC 69  
 Db 1 GSCGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGCTGGTGTTCCTCGGGGAGCAAGC 60  
 QY 70 GTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGCTTTTGACACACCTATCAG 129  
 Db 61 GTTGAAAGACATCTTTGATCACCAGATTCAGGTATGACAGCTTTTGACACACCTATCAG 120  
 QY 130 GCAATAATTGGCATTGACCTTTTATCAAAAACATATGTACTTGGAGGTGGAACAATCGGG 189  
 Db 121 GCAACAATTGGCATTGACCTTTTATCAAAAACATATGTACTTGGAGGTGGAACAATCAGG 180  
 QY 190 CTTCCGCTGTGGGATACGGCGGTCAGGAACGTCCTCCGTAGCCTCATTTCCAGGTACATC 249  
 Db 181 CTTCCAGCTGTGGGATACGCGGTCAGGAACGTCCTCCGTAGCCTCATTTCCAGGTACATC 240  
 QY 250 CGTGATTCTGCTGACGCTGTAGTGTACAGATATCACAATGTTAACTCATTCAGCANA 309  
 Db 241 CGTGATTCTGCTGACGCTGTAGTGTACAGATATCACAATGTTAACTCATTCAGCANA 300  
 QY 310 ACTACAAAGTGGATTGATGTCTAGACAGAAAGAGGAAGTGTATGTTATCATCATGCTA 369  
 Db 301 ACTACAAAGTGGATTGATGTCTAGACAGAAAGAGGAAGTGTATGTTATCATCATGCTA 360  
 QY 370 GTAGGAAATAGAACAGATCTTCTGCACAGAGGCAAGTGTCTAGTTGAGGAGGAGAGG 429  
 Db 361 GTAGGAAATAGAACAGATCTTCTGCACAGAGGCAAGTGTCTAGTTGAGGAGGAGAGG 420  
 QY 430 -AAGCCAAAGGCTGAATGTTACGTTTATGAACTAGGCGCAAACTGGATACAATGTA 489  
 Db 421 AAAGCCAAAGAGCTGAATGTTATGTTTATGAAACTAGTGCAAAAGCTGGATACAATGTA 480  
 QY 490 AAGCAGCTCTTTCGACGCTAGCAGCAGCTTTCCGCGGAATGGAAGACACACAGACGGA 549  
 Db 481 AAGCAGCTCTTTCGACGCTAGCAGCAGCTTTCCGCGGAATGGAAGACACACAGACAGA 540  
 QY 550 AGCAGAGAAGACATGAGTGACATATAAACTGGAAAAGCCTCAGGAGCAAACTCAGCGAA 609  
 Db 541 AGCAGAGAAGATATGATTGACATATAAACTGGAAAAGCCTCAGGAGCAAACTCAGTGAA 600  
 QY 610 GGGGTTGTTCTCTGC 624  
 Db 601 AGAGGCTGTCTCTGC 615

## RESULT 8

BQ632282  
 LOCUS  
 DEFINITION  
 BQ632282  
 i125a01.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6030984 3'  
 similar to SW:RAB6\_HUMAN P20340 RAS-RELATED PROTEIN RAB-6. [1] ;  
 mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 609)  
 REFERENCE  
 AUTHORS  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y., and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)

# TITLE JOURNAL COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Possible reversed clone: similarity on wrong strand

Seq primer: -400P from Gibco

High quality sequence stop: 439.

Location/Qualifiers

1..609

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6030984"

/tissue\_type="Purified pancreatic islet"

/lab\_host="DH10B"

/clone\_lib="HR85 islet"

/note="organ: Pancreas; Vector: pBluescript SK(-); Site\_1: NotI; Site\_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

## ORIGIN

Query Match 74.9%; Score 570.6; DB 5; Length 609;  
Best Local Similarity 96.1%; Pred No. 6.2e-160;  
Matches 585; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

QY 16 GACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTTCCTGGGGGACCAAGCGTTGCA 75
   |||||
Db 1 GACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTTCCTGGGGGACCAAGCGTTGGA 60
   |||||
QY 76 AAGACATCTTTGATCACCAGATTACAGGTATGACAGTTTGTGACACACCTATCAGGCAATA 135
   |||||
Db 61 AAGACATCTTTGATCACCAGATTACAGGTATGACAGTTTGTGACACACCTATCAGGCAACA 120
   |||||
QY 136 ATTGCGATTGACTTTTATCAAAAATATGACTTGGAGGATGGAACAATCGGCGTTGG 195
   |||||
Db 121 ATTGCGATTGACTTTTATCAAAAATATGACTTGGAGGATGGAACAATCAGGCTTCAG 180
   |||||
QY 196 CTGTGGGATACGGCGGGTCAGGAACGTCCTCGTAGCCCTCAITTCACAGGTACATCCGTGAT 255
   |||||
Db 181 CTGTGGGATACGCGGGTCAGGAACGTTTCGCTAGCCTCAITTCACAGTTACATCCGTGAT 240
   |||||
QY 256 TCTGCTGAGCTGTAGTAGTTTACATATACAAATGTTAACTATTCACGAACATACA 315
   |||||
Db 241 TCTGCTGAGCTGTAGTAGTTTACATATACAAATGTTAACTATTCACGAACATACA 300
   |||||
QY 316 AAGTGGATTGATGTCCAGAACAGAGGAAGTGAATGTTATCATCACCTAGTAGGA 375
   |||||
Db 301 AAGTGGATTGATGTCCAGAACAGAGGAAGTGAATGTTATCATCATCTAGTAGGA 360
   |||||
QY 376 AATGAACAGATCTTGTCTGACAAAGGCAAGTGTCAAGTGGAGGGGAGAGGAAAGCC 435
   |||||
Db 361 AATAAACAAGATCTTGTCTGACAAAGGCAAGTGTCAATTTGAGGGGAGAGGAAAGCC 420
   |||||

```

```

QY 436 AAAGGGCTGAATGTTACGTTTATTGAAACTAGGCGCAAAAACTGGATACAAATGTTAAAGCAG 495
   |||||
Db 421 AAAGAGCTGAATGTTATGTTTATTGAAACTAGTGCAAAAGCTGGATACAAATGTTAAAGCAG 480
   |||||
QY 496 CTCCTTTCCAGCTGTAGCAGCAGCTTTTCCCGGGAATGAAAGCACACAGGACGGAAGCAGA 555
   |||||
Db 481 CTCCTTTCCAGCTGTAGCAGCAGCTTTTCCCGGGAATGAAAGCACACAGGACGGAAGCAGA 540
   |||||
QY 556 GAAGACATGAGTGACATAAACTGGAAGAGCTCAGGAGCAAAACAGTCAGGAAAGGGGT 615
   |||||
Db 541 GAAGATATGATTGACATAAACTGGAAGAGCTCAGGAGCAAAACAGTCAGTGAAGAGGC 600
   |||||
QY 616 TGTTCCTGTC 624
   |||||
Db 601 TGTTCCTGTC 609
   |||||

```

## RESULT 9

BI694279

LOCUS

DEFINITION

603347724F1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5375477 5', mRNA sequence.

ACCSSION

BI694279

VERSION

BI694279.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 790)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-x@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM11955 row: p column: 06

High quality sequence stop: 788.

Location/Qualifiers

1..790

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N-3"

/db\_xref="taxon:10090"

/clone="IMAGE:5375477"

/tissue\_type="tumor, biopsy sample"

/dev\_stage="5 months"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Mam2"

/notes="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match 74.3%; Score 566; DB 4; Length 790;

Best Local Similarity 85.9%; Pred. No. 1.6e-158;

Matches 656; Conservative 0; Mismatches 95; Indels 13; Gaps 2;

QY 11 GCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGT-GTTCCTGGGGGACCAAGC 69

Db 1 GCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGGTTCCTGGGAGACGAGC 60

QY 70 GTTGCAAGAAGCATCTTTTGATCACAGATTACAGGTATGACAGTTTGTGACACCTATCAG 129

Db 61 GTTGCAAGAAGCATCTTTTGATCACAGATTACAGGTATGACAGTTTGTGACACCTATCAG 120





Query Match 69.3%; Score 528; DB 6; Length 744;  
Best Local Similarity 85.7%; Pred. No. 4.3e-147;  
Matches 615; Conservative 0; Mismatches 90; Indels 13; Gaps 2;

QY 57 GGGGAGCAAGCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTGA 116  
DB 3 GGGAGAGCAGAGCGTTGGAAGACGCTCTTGATCACCAGATTCATGATGACAGTTTGA 62  
QY 117 CAACACCTATCAGGCAATATTTGGCATTCAGCTTTTATCAAAACCTATGATCTTGAGGA 176  
DB 63 CAACACCTATCAGGCAATATTTGGCATTCAGCTTTTATCAAAACCTATGATCTTGAGGA 122  
QY 177 TGGAAACATCGGCTTCGGCTGTGGATACGGGGTCAGGACGCTCCGTCAGCTCAT 236  
DB 123 TAGAACCGTGCCTGATTCGAATTTATGGACACAGCAGGTCAAGAGCGTTTCAGGAGCTTGT 182  
QY 237 TCCAGGTACATCCGTTGATTCGTCAGCTGTAGTATGTTTACGATATCACAAATGTTAA 296  
DB 183 TCTAGCTACATTCGAGACTCCACTGTGGCAGTGTGTTTATGATATCACAAATGTTAA 242  
QY 297 CTCAATTCAGCAAACTACAAAGTGGATTCAGATGATGATGATGATGATGATGATGATGAT 356  
DB 243 CTCAATTCAGCAAACTACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 302  
QY 357 TATCATCAGCTAGTAGGAAATAGAACAGATCTTGCTGACAGAGCAAGTGTCAAGTTGA 416  
DB 303 CATCATCATGCTAGTAGGAAATAGAACAGATCTTGCTGATAGAGCAAGTGTCAATTTGA 362  
QY 417 GGAGGGAGAGAGAAAGCCAAAGGCTGAATGTTACGTTTATTTGAACTAGGGCAAAAC 476  
DB 363 GGAGGGAGAGAGAAAGCCAAAGAGCTGAATGTTATGTTTATTTGAAACCAAGTGCAAAAGC 422  
QY 477 TGGATACAAATGTAAGAGCCTCTTTTCAGCTGTAGCAGCAGCTTTTCGGGGAAATGGAAG 536  
DB 423 AGGATACAAATGTAAGAGCCTTTTCGGGCTGTGTCTGAGCTTTTACCTGGAATGGAAG 482  
QY 537 CACACAGGACGGAAGCAGAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 596  
DB 483 CACACAGGACGGAAGCAGAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 542  
QY 597 AACAGTCAGCAAGGGGTTGTTCTGCTACTCTCCATGTCATCTTCAACCTTCCTCA 656  
DB 543 ACCAGTCAATGAAGCGGCTGTCTGCTGAATTCCTCCCTTGGCATCTTTCAACCTTCCTCA 602  
QY 657 GAA-----GCCCTTACTCTTTCACTGACCTGACGTGGAATATTGGCTTGA 704  
DB 603 GAACTCACTGCTTTGGCCCCAFACTCTTTCAATGACTGCAATGGAATATTGGCTTGA 662  
QY 705 CC-TTTTCCCTTCATTAATACGTTTTCGAATTCATCATTTGCTGCTGTCTCGTGAG 761  
DB 663 CTTTTTCCCTTCAGTAATACGTTATGCAATTCATCATTTGCTGCTGTCTCGTGAG 720

RESULT 13  
BE965054  
LOCUS  
DEFINITION BE965054 762 bp mRNA linear EST 14-DEC-2000  
6016589391 NTH\_MGC\_69 Homo sapiens cDNA clone IMAGE:3886281 3',  
mRNA sequence.  
ACCESSION BE965054  
VERSION BE965054.2 GI:11769028  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 762)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT On Oct 3, 2000 this sequence version replaced gi:10575759.  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTP/Gazdar

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLC649 row: n column: 10  
High quality sequence start: 43  
High quality sequence stop: 728.  
Location/Qualifiers  
1..762  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3886281"  
/tissue\_type="large cell carcinoma, undifferentiated"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_69"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."

FEATURES  
source

## ORIGIN

Query Match 68.9%; Score 525.2; DB 2; Length 762;  
Best Local Similarity 87.4%; Pred. No. 3e-146;  
Matches 626; Conservative 0; Mismatches 73; Indels 17; Gaps 4;

QY 22 GGGATCCGCTGAGGAAATTCAGCTGTGTTCCTGGGGGAGCAAGGTTGCAAGACA 81  
DB 39 GTGAATCCGCTGAGGAAATTCAGCTGTGTACCTGGGGGAGCAAGGTTTGGAAAGACA 98  
QY 82 TCTTTGATCACACAGATTCAGGTATGACAGTTTTCACACACCTATCAGGCAATATTGSC 141  
DB 99 TCTTTGATCACACAGATTCAGGTATGACAGTTTTCACACACCTATCAGGCAATATTGSC 158  
QY 142 ATTGACTTTTATCAAAACCTATGTA--CTTGGAGGATGGA---ACAATCGGGCTTCGGC 196  
DB 159 ATTGACTTTTATCAAAACCTATGTAGCTTTGGAGGATCGACACAGTAGCGAGTTGCAAT 218  
QY 197 TGTGGGATACGGGGGTCAGGAAAGTCTCCGTAGCTCATTTCCAGGTACATCCGTGAT 256  
DB 219 TATGGGACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTGATTCCTAGCTACATTCGTGACT 278  
QY 257 CTGCTGCAGCTGTAGTATGTTTACGATATCACAATATGTTAACTCATTTCCAGCAAACTACAA 316  
DB 279 CCACGTGGCAGTTGTTGTTTATGATATCACAATATGTTAACTCATTTCCAGCAAACTACAA 338  
QY 317 AGTGGATGATGATGTACAGAACAGAGGAAAGTGTATATCATCATCGCTAGTAGGAA 376  
DB 339 AGTGGATGATGATGTACAGAACAGAGGAAAGTGTATATCATCATCTAGTAGGAA 398  
QY 377 ATAGAACAGATCTTGTGACAGAGGCAAGTGTCTAGTTGAGGAGGAGAGGAAAGCCA 436  
DB 399 ATAAACACAGATCTTGTCTGACAGAGGCAAGTGTCAATTTGAGGAGGAGAGGAAAGCCA 458  
QY 437 AAGGCTCAATGTTACGTTTATGAAACTAGGCAAAACCTGGATACAAATGTAAAGCAGC 496  
DB 459 AAGAGCTGAATGTTATGTTTATGAACTAGTGCNAAAGCTGGATACAAATGTAAAGCAGC 518  
QY 497 TCTTTCCAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGCACACAGAGCGGAAGCAGAG 556  
DB 519 TCTTTCCAGCTGTAGCAGCAGCTTTTCCGGGAATGGAAGCACACAGAGCGGAAGCAGAG 578  
QY 557 AAGACATGAGTGACATAAACTGGAAGAGCTCAGGAGCAAACTAGTCAGCGAAGGGGTT 616  
DB 579 AAGATATGATTGACATAAACTGGAAGAGCTCAGGAGCAACCACTAGTGAAGAGGCT 638  
QY 617 GTTCTGCTACTCTCCCATGTCTTCAACC-----CTTCTCTCAGAGCCGCC 665  
DB 639 GTTCTGCTAACTCTCCCATGTCTTCAACCTTCTTTCAGAGGCTCACTGCTTTGGCCCC 698  
QY 666 TTACTCTTTTCATCTGACTGAGTGAATATTTGGCTTGAACCTTTTCCCTTTCATTAA 721



Db 699 TTAATCTTTTCATTGACTGC-GTGTGACTATTGGTGTGAACCTTTTCCTTCATAATAA 753

|||||

BF431157 632 bp mRNA linear EST 29-NOV-2000  
7008c05.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:3573489 3'  
similar to SW:RAB6\_HUMAN P20340 RAS-RELATED PROTEIN RAB-6. [1] ;  
mRNA sequence.

ACCESSION BF431157  
VERSION BF431157.1 GI:11443271  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 632)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
[info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -40UP from Gibco  
High quality sequence scop: 494.  
Location/Qualifiers  
1..632  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3573489"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Kid11"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 132376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

ORIGIN  
Query Match 68.6%; Score 522.6; DB 2; Length 632;  
Best Local Similarity 91.1%; Pred No. 1.7e-145;  
Matches 555; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATGTCGCGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGTGGTCTCTCGGG 60  
Db 24 ATGTCCACGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGAGTGGTCTCTCGGG 83

Qy 61 GAGCAACGGTTGCAACACATCTTGTATCACCAGATTTCAGGTATGACAGTTTTCACAAAC 120  
Db 84 GAGCAACGGTTGGAAGACATCTTTGTATCACCAGATTTCATGTATGACAGTTTTCACAAAC 143

Qy 121 ACCTATCAGGCAATAATTTGGCATTGACATTTTATCATAAAATATGCTACTTCGAGGATCGA 180  
Db 144 ACCTATCAGGCAACAAATTTGGCATTGACATTTTATCATAAAATATGCTACTTCGAGGATCGA 203

Qy 181 ACAATCGGGCTTCGGCTGTGGATACGCGGGTCAGGAACTGTCCTCGGAGCTCATTCCTCC 240  
Db 204 ACAGTACGATTGCAATTATGGAACACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTCGATTCCT 263

Qy 241 AGGTACATCCGTGATTCTGCTGCAGCTGTACTAGTTTACGATATCACAAATGTTAACTCA 300  
Db 264 AGCTACATTCGTGACTCCACTGTGGCAGTTGTTGTTTATGATATACAAATGTTAACTCA 323

Qy 301 TTCCAGCAAACTACAAAGTGGATTGATGTGAGAACAGAAAGAGGAAGTGATGTTATC 360  
Db 324 TTCCAGCAAACTACAAAGTGGATTGATGTGAGAACAGAAAGAGGAAGTGATGTTATC 383

Qy 361 ATCAGCGTAGTAGAAATAGAACAGATCTTCTGACAAAGAGGCAAGTGTCTAGTTGAGGAG 420  
Db 384 ATCATGTCTAGTAGAAATATAAACAGATCTTCTGACAAAGAGGCAAGTGTCTCAATTGAGGAG 443

Qy 421 GGAGAGGGAAGCCAAAGGGCTCAATGTTTATTTAGAACTAGGCGCAAAACTCGA 480  
Db 444 GGAGAGGGAAGCCAAAGAGCTGAATGTTTATTTGAAACTAGTGTGAAAAGCTCGA 503

Qy 481 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGGAATCGAAAGCACA 540  
Db 504 TACAATGTAAAGCAGCTCTCTCGACGTGTAGCAGCAGCTTTGCGGGGAATCGAAAGCACA 563

Qy 541 CAGGACGGAAGCAGAGAACATGAGTGACATATAAACTGGAAAAAGCTTCAGGAGCAAA 600  
Db 564 CAGGACGGAAGCAGAGAACATGATGATTGACATATAAACTGGATAAGCCCTCATGAGCAACA 623

Qy 601 GTCAGCGAA 609  
Db 624 GTCAGTGAA 632

RESULT 15  
LOCUS CA316338  
DEFINITION UI-M-FW0-cbp-i-24-0-UI.r1 NIH\_BMAP\_FW0 Mus musculus cDNA clone  
IMAGE:6812737 5', mRNA sequence.  
ACCESSION CA316338  
VERSION CA316338.1 GI:24534462  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 792)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.  
Location/Qualifiers  
1..792  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6812737"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_FW0"  
/notes="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose

FEATURES  
source



gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is ACGGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match		67.2%;	Score 512;	DB 6;	Length 792;
Best Local Similarity		88.1%;	Pred. No. 2.8e-142;		
Matches 557;		Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0;
Qy	1	ATGTCGGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTC	AAAGCTGGTGTCTCTGGGG	60	
Db	65	ATGTCGGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTC	AAAGCTGGTGTCTCTGGGA	124	
Qy	61	GAGCAAGCGTTGCCAAGACATCTTTGATCACCAGATTCCAGGTATGACAGTTTTCGACAC	120		
Db	125	GAGCAGAGCGTTGGAAAGACGCTCTTGATCACCCTGATTCATGTATGACAGTTTTCGACAC	184		
Qy	121	ACCTATCAGGCAATTAATTCGGCAATTCGACTTTTATCAAAAATCTATGTACTTGGAGGATGGA	180		
Db	185	ACCTATCAGGCAATTAATTCGGCAATTCGACTTTTATCAAAAATCTATGTACTTGGAGGATGGA	244		
Qy	181	ACAATCGGCTTCGGCTGTGGATACGGCGGTGAGGAACGTCCTCGTAGCCTCAATTC	240		
Db	245	ACCGTCCGATTGCAATTTATGGGACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTGATTCT	304		
Qy	241	AGGTACATCCGTGATTCTGCTGACGTGTAGTGTACGATATCACAAATGTTAACTCA	300		
Db	305	ACCTACATTCGAGACTCCACTGTGGCAGTTGTTTATGATATCACAAATGTTAACTCA	364		
Qy	301	TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAAACAGAAAGAGGAGTGTATC	360		
Db	365	TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAAACAGAAAGAGGAGTGTATC	424		
Qy	361	ATCAGCTAGTAGGAAATAGAACAGATCTTGTGACAGAGGCAAGTGTCACTTGAAGGAG	420		
Db	425	ATCATGCTAGTAGGAAATAGAACAGATCTTGTGATAGAGGCAAGTGTCACTTGAAGGAG	484		
Qy	421	GGAGAGAGAAAGCCAAAGGCTGAATGTTAGCTTTATTGAACTAGGCAAAACTGGA	480		
Db	485	GGAGAGAGAAAGCCAAAGGCTGAATGTTAGCTTTATTGAACTAGGCAAAAGCAGGA	544		
Qy	481	TACAATGTAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGGGGAATGGAAGCACA	540		
Db	545	TACAATGTAAGCAGCTTTTCGCGCTGTGCTGACGCTTTTACCTGGATGGAAGCACA	604		
Qy	541	CAGGACGGAAGCAGAGACATAGTGTACATAAACTGGAAGAGCCTCAGGAGCAACA	600		
Db	605	CAGGACGGAAGCAGAGACATAGTGTACATAAACTGGAAGAGCCTCAGGAGCAACA	664		
Qy	601	GTCAGGAGAGGGGTTGTTCTCTACTCTCC	632		
Db	665	GTCATGAAAGCGGCTGTTCTGCTAATCCCC	696		

Search completed: April 25, 2005, 11:40:30  
Job time : 3105 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2005, 22:22:10 ; Search time 3301 Seconds  
(without alignments)  
3728.453 Million cell updates/sec

Title: US-09-441-857-12  
Perfect score: 1304  
Sequence: 1 MSAGGDFGNLKRKFLVFLG.....NLFPSTLTFNCSSLLIPVSWR 254

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame\_plus\_p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US09441857/runat\_22042005\_155508\_24747/app\_query.fasta\_1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09441857 @CGN 1 1 5600 @runat\_22042005\_155508\_24747 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htc.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sv.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1304	100.0	765	9 CR533469 Homo sapi
2	1304	100.0	840	9 AF309646 Homo sapi
3	1304	100.0	3084	6 AX086357 Sequence
4	1299	99.6	198961	9 AC079776 Homo sapi

C	5	1299	99.6	349981	6	AX647880	Sequence
C	6	1299	99.6	349989	6	AX647177	Sequence
C	7	1293	99.2	165411	9	AC018865	Homo sapi
C	8	1288	98.8	195280	9	AC073869	Homo sapi
C	9	1261	96.7	242865	2	AC120839	Pan trogl
	10	1255	96.2	184706	9	AC120781	Homo sapi
	11	1147	88.0	3070	9	BC003617	Homo sapi
	12	1147	88.0	3100	9	BC068486	Homo sapi
	13	1147	88.0	3120	9	AK057157	Homo sapi
	14	1138	87.3	3745	6	AB447780	Sequence
	15	1137	87.2	1770	10	BC019118	Mus muscu
	16	1089	83.5	1950	10	AB041575	Mus muscu
	17	1062.5	81.5	1170	5	CR390639	Gallus ga
	18	1012.5	77.6	844	5	AF540409	Melagris
	19	1000	76.7	2527	10	BC026915	Mus muscu
	20	982	75.3	739	9	AF119836	Homo sapi
	21	960	73.6	683	9	AF130122	Homo sapi
	22	959	73.5	740	6	AX771444	Sequence
	23	959	73.5	740	9	HUMRAB6A	Sequence
	24	959	73.5	2099	5	BC046683	Xenopus l
	25	953	73.1	627	9	AF198616	Homo sapi
	26	953	73.1	627	9	AF498941	Homo sapi
	27	948	72.7	627	9	CR541984	Homo sapi
	28	946.5	72.6	1461	5	BC064230	Xenopus t
	29	944	72.4	627	9	AF130986	Homo sapi
	30	944	72.4	627	9	AF498939	Homo sapi
	31	897.5	68.8	4768	10	BC060618	Mus muscu
	32	892	68.4	627	9	AF498940	Homo sapi
	33	892	68.4	627	9	BT007263	Homo sapi
	34	892	68.4	1261	9	BC002510	Homo sapi
	35	892	68.4	1266	9	AF166492	Homo sapi
	36	888	68.1	5395	9	BC078662	Homo sapi
	37	884	67.8	1589	6	BD131169	Human tra
	38	883	67.7	573	9	AF124200	Homo sapi
	39	878	67.3	1652	5	BC074238	Xenopus l
	40	869	66.6	2005	5	BC080215	Danio rer
	41	864.5	66.3	172962	10	AL713865	AL713865 Mouse DNA
	42	860	66.0	2498	5	BC058059	BC058059 Danio rer
	43	846.5	64.9	1157	3	D84314	D84314 Drosophila
	44	846.5	64.9	1188	6	C0586572	C0586572 Sequence
	45	846.5	64.9	1857	3	AY060261	AY060261 Drosophil

## ALIGNMENTS

RESULT 1	CR533469	765 bp	mRNA	linear	PRI 22-JUN-2004
LOCUS	CR533469	Homo sapiens full open reading frame cDNA clone RZPD0834C0517D for gene RAB6C, RAB6C, member RAS oncogene family; complete cds, incl. stopcodon.			
DEFINITION	CR533469.1	GI:49065363 Full ORF shuttle clone, Gateway(TM), complete cds.			
ACCESSION	CR533469	Homo sapiens (human)			
VERSION	CR533469.1	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
KEYWORDS	1 (bases 1 to 765)				
SOURCE	Ebert.L., Schick.M., Neubert.P., Schatten.R., Henze,S. and Korn.B.				
ORGANISM	Cloning of human full open reading frames in Gateway(TM) system entry vector (pDONR201)				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 765)				
TITLE	Ebert.L., Schick.M., Neubert.P., Schatten.R., Henze,S. and Korn.B.				
JOURNAL	Submitted (21-JUN-2004)				
REFERENCE	Submitted (21-JUN-2004)				
AUTHORS	Submitted (21-JUN-2004)				
TITLE	Submitted (21-JUN-2004)				
JOURNAL	Submitted (21-JUN-2004)				
COMMENT	Submitted (21-JUN-2004)				

www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834  
 www.rzpd.de/products/orfclones/  
 Contact: Inge Ariart  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 100  
 Fax: +49 30 32639 111  
 www.rzpd.de

This clone is available from RZPD;  
 contact RZPD (customer.service@rzpd.de) for further information.  
 This CDS clone is a part of a collection of human full ORF clones  
 generated by RZPD.  
 This CDS has been cloned incl. stopcodon.  
 The CDS has been inserted into pDONR201 via a BP Clonase(TM)  
 reaction. Additional sequence has been added in front of the start  
 codon: att..AAAAA GCA GGC (ATG).  
 After the stop codon 3' UTR sequence is present in front of the 3'  
 att site (ACCAGCTTTCTT).  
 Compared to the reference sequence NM\_032144 (gi14149798) we did  
 not find any amino acid exchanges.  
 Clone distribution: http://www.rzpd.de/products/orfclones/.

## FEATURES

source

```

1..765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="RZPD0834C0517D"
/clone_lib="Human Full ORF Clones Gateway(TM) - RZPD"
/lab_host="DH10B"
/notes="Vector: pDONR201, Site_1: attP1; Site_2: attP2"
1..765
/gene="RAB6C"
1..765
/gene="RAB6C"
/codon_start=1
/protein_id="CAG38500.1"
/db_xref="GI:49065364"
/translation="MSAGDFGNPLKFKLYFLGEQSVAKTSLITFRYDSFDNTYQA
IIGDIFSKTMYLEDGTIGRLMDTAGQRLRSILIPYIRDSAAAAYVVDITNVNSFQ
QTTKWDVTRGSDVITLVGNRTDLADKQVSEVEGERKAKGLNVTFIETRAKTG
YNVKQLFRVAAALPGMESTQDGSREDMSDIKLEKPEQTVSEGGSCSYSPMSSSTLP
OKPPYSFIDCSVNIGLNLFPSSLTPVSWR"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 4,1e-118 Length: 765
Score: 1304.00 Matches: 254
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

```

US-09-441-857-12 (1-254) x CR533469 (1-765)

```

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 1 ATGTCGCGGGGAGACATTCGGGAATCCGTCGAGGAATTCAGAGTGTCTCTCGGG 60

Qy 21 GluInSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40
Db 61 GAGCAAGCGTTCGCAAGACATCTTGTATCACCAGATTCAGGTATGACAGTTTTGACAAC 120

Qy 41 ThrTyrGlnAlaIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 121 ACCTATCAGGCAATTAATGGCAATTCGACATTTTATCAAAACATATGTCTTGAGAGTGA 180

Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 181 ACAATCGGCTTCGGCTGTGGATACCGCGGGTCAGAACGTCCTCCGTAGCCTCATTC 240

Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 241 AGGTACATCCGTGATTCGTCGACGCTGTAGTAGTTTACGATATACAAATGTTAACTCA 300

```

```

Qy 101 PheGlnGlnThrThrLysTyrPheIleAspAspValArgThrGluArgGlySerAspValle 120
Db 301 TTCCAGCAAACTACAAAGTGGATTGATGATCTCAGAACAGAAAGAGGAAGTGTATGTTATC 360

Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 361 ATCAGCTAGTAGGAATATAGAACAGATCTTCTGCAAGAGGCAAGTGTCAAGTTGAGAG 420

Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 421 GGAGAGAGGAAGCAAAAGGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAACTGGA 480

Qy 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLalaLeuPheProGlyMetGluSerThr 180
Db 481 TACAATGTAAGACAGCTCTTTTCGACGCTGTAGCAGCAGCTTTGCCGGGAATGGAAGACA 540

Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 541 CAGGACGAGCAGAGAGACATGATGATGACATATAAACTGGAAAAGCCTTCAGAGCAACA 600

Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
Db 601 GTCAGCGAAGGGGTTGTTCTCTGCTACTCTCCCATGTCATCTTCAACCTTCTCAGAG 660

Qy 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240
Db 661 CCCCCTTACTCTTTTCAATGACTGCAGTGTGAATATATGCTTGAACCTTTTCCCTTCATTA 720

Qy 241 IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254
Db 721 ATAACGTTTTGCAATTCATTCCTGCTGCTCTCTCGTGAGG 762

```

## RESULT 2

AF309646

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

1..840

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2q31"

48..812

/notes="similar to GTP-binding protein"

/codon\_start=1

/product="Rab6-like protein WTH3"

/protein\_id="AA039851.1"

/db\_xref="GI:23820936"

/translation="MSAGDFGNPLKFKLYFLGEQSVAKTSLITFRYDSFDNTYQA

IIGDIFSKTMYLEDGTIGRLMDTAGQRLRSILIPYIRDSAAAAYVVDITNVNSFQ

QTTKWDVTRGSDVITLVGNRTDLADKQVSEVEGERKAKGLNVTFIETRAKTG

YNVKQLFRVAAALPGMESTQDGSREDMSDIKLEKPEQTVSEGGSCSYSPMSSSTLP

ORGANISM

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 REFERENCE  
 AUTHORS Wiemann, S.  
 TITLE Human dna sequences  
 JOURNAL Patent: WO 0112659-A 309 22-FEB-2001;  
 German Human Genome Project (DE)  
 FEATURES  
 source Location/Qualifiers  
 1..3084  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,58e-117 Length: 3084  
 Score: 1304.00 Matches: 254  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-09-441-857-12 (1-254) x AX086357 (1-3084)  
 QY 1 MetSerAlaGlyClyAspPheGlyAanProLeuArgLysPheLysLeuValPheLeuGly 20  
 Db 456 ATGTCCGGCGGGAGACATTCGGGAATCCGTCAGGAAATCAAGCTGGTGTCTCTCGGG 515  
 QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAan 40  
 Db 516 GAGCAAGCGTTGCAAGACATCTTTGATCACCAGATTTCAGGTATGACAGTTTGACAAC 575  
 QY 41 ThrTyrGlnAlaIleIleGlyLeAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
 Db 576 ACCTATCAGGCAATAATGGCATTTGACTTTTATCAAAACATATGACTTGGAGGATGGA 635  
 QY 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
 Db 636 ACNATCGGGCTTCGGCTGGGATACGGGGTCAGAGCGTCTCCGTACGCTCATTTCCC 695  
 QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAenValAenSer 100  
 Db 696 AGGTACATCCGTGATTCTCTGCAGCTGTAGTAGTTTACGATATCAAAATGTTAACTCA 755  
 QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120  
 Db 756 TTCCAGCAAACTACAAGTGGATTGATGTGAGAACAGAAAGAGAGAACTGATGTATC 815  
 QY 121 IleThrLeuValGlyValAanArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
 Db 816 ATCACCTAGTAGGAATAGAACAGATCTTCTGACACAGAGCGCAAGTGTCTGAGGAG 875  
 QY 141 GlyGluArgLysAlaLysGlyLeuAanValThrPheIleGluThrArgAlaLysThrGly 160  
 Db 876 GGAGAGAGAAAGCCAAAGGCTGAATGTTACGTTTTATTGAAACTAGGCGCAAAACTGGA 935  
 QY 161 TyrAanValLysGlnLeuPheArgArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
 Db 936 TACAATGTAAAGCAGCTCTTTCCAGCTGTAGCAGCAGCTTTGCCGGGAATGGAAGACCA 995  
 QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
 Db 996 CAGGACGGAGCAGAGAGACATGAGTGACATNAAACTGGAAAAAGCTCAGGAGCAAAACA 1055  
 QY 201 ValSerGluGlyClyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220  
 Db 1056 GTCAGCGAAAGGGGTGTTCTCTGCTACTCTCCCATGTCTTCAACCCCTTCCTCAGAAG 1115  
 QY 221 ProProTyrSerPheIleAspCysSerValAanIleGlyLeuAanLeuPheProSerLeu 240  
 Db 1116 CCCCTTACTCTTTCACTGACGTGTAATTTGGCTTGAACCTTTTCCCTTCATTA 1175

Qy 241 lletThrPheCysAsnSerSerLeuLeuProValSerTtpAtg 254  
 |||||  
 Db 1176 ATACGTTTTCATCATCATCTGCTGCTGCTGAGG 1217

RESULT 4  
 AC079776  
 LOCUS 198961 bp DNA linear PRI 22-FEB-2002  
 DEFINITION Homo sapiens BAC clone RP11-315H12 from 2, complete sequence.  
 ACCESSION AC079776  
 VERSION AC079776.5 GI:18482321  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 198961)  
 Sultston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 198961)  
 AUTHORS Shah,N. and Kozlowicz,A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-315H12  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 198961)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 198961)  
 REFERENCE 4 (bases 1 to 198961)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 5 (bases 1 to 198961)  
 REFERENCE 5 (bases 1 to 198961)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 6 (bases 1 to 198961)  
 REFERENCE 6 (bases 1 to 198961)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Feb 3, 2002 this sequence version replaced gi:18042404.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0315H12  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa K., Woon, P.Y., Zhao B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-228P15, 2000 bp overlap; the clone sequenced to the right is RP11-32C20. Actual start of this clone is at base position 129432 of RP11-228P15; actual end is at base position 198961 of RP11-315H12.

Data from AC079450 was used to finish AC079776.

FEATURES	Location/Qualifiers
source	1..198961
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-315H12"
	/clone_lib="RPCI-11"
repeat_region	216..817
	/rpt_family="L1"
repeat_region	818..1399
	/rpt_family="MaLR"
repeat_region	1403..2150
	/rpt_family="MaLR"
repeat_region	2148..2492
	/rpt_family="L1"
repeat_region	2534..3082
	/rpt_family="L1"
repeat_region	3106..3742
	/rpt_family="MaLR"
repeat_region	3763..4054
	/rpt_family="Alu"
repeat_region	4084..4452
	/rpt_family="MaLR"
repeat_region	4456..4666
	/rpt_family="Alu"
repeat_region	4671..4753
	/rpt_family="L1"
repeat_region	4739..5052
	/rpt_family="L1"
repeat_region	5057..5148
	/rpt_family="L1"
repeat_region	5149..5262
	/rpt_family="Alu"
repeat_region	6702..6877
	/rpt_family="Alu"
repeat_region	6878..7812
	/rpt_family="L1"
repeat_region	7812..8134
	/rpt_family="L1"
repeat_region	8135..8434
	/rpt_family="Alu"
repeat_region	8435..9553
	/rpt_family="L1"
repeat_region	9554..9574
	/rpt_family="AT-rich"
repeat_region	9579..9658
	/rpt_family="L1"
repeat_region	9659..9961
	/rpt_family="Alu"
misc_feature	9936..9958



```
/note="old seq 1369: original length 744.802 splitted into
3 new seqs-new seq 1369: from 000.001 to 349.980-new seq
2072: from 300.001 to 649.980-new seq 2073: from 600.001
to 744.802"
```

## ORIGIN

Alignment Scores:	4.11e-114	Length:	349981
Pred. No.:	1299.00	Matches:	253
Score:	1299.00	Conservative:	0
Percent Similarity:	99.61%	Mismatches:	1
Best Local Similarity:	99.61%	Indels:	0
Query Match:	99.62%	Gaps:	0
DB:	6		

US-09-441-857-12 (1-254) x AX647880 (1-349981)

Qy	1	MetSerAlaGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly	20
Db	45699	ATGTCGCGCGCGGAGACTTCGGGAAATCCGCTGAGGAAATCAAGCTGTGTCTCGGG	45640
Qy	21	GlulnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn	40
Db	45639	GAGCAAGCGTTGCAAGACATCTTTGATCACAGATTGAGGTATGACAGTTTTGACAA	45580
Qy	41	ThrTyrGlnAlaIleIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly	60
Db	45579	ACCTATCAGGCAATAATTGGCATTGACTTTTATCAAAACATATGTACTTGGAGGATGA	45520
Qy	61	ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro	80
Db	45519	ACAATCGGGCTTCGGCTGTGGGATACGGCGGGTCAGGAACTGTCCTCGTAGCCTCATTTCC	45460
Qy	81	ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer	100
Db	45459	AGGTACATCGTGATTCTCTCGAGCTGTAGTAGTTTACGATATCACAAATGTTAACTCA	45400
Qy	101	PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle	120
Db	45399	TTCCAGCAAACTACAAAGTGGATTGATGTGCAGAACAGAAAGGAAGTGATGTTATC	45340
Qy	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu	140
Db	45339	ATCACGCTAGTAGGAAATAGAACAGATCTTGTCTGACAGAGAGCAAGTGTCAAGTTCAGAG	45280
Qy	141	GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly	160
Db	45279	GGAGAGAGGAAGCCAAAGGCTGAATGTTAGTTTATTGAAACTAGGGCAAAAGCTGGA	45220
Qy	161	TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr	180
Db	45219	TACAAATGTAAGCAGCTCTTTCGACGTGTAGCAGCAGCTTTTCGCGGAATGGAAGCAC	45160
Qy	181	GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluInThr	200
Db	45159	CAGGACGGAAGCAGAGAAGACATGAGTGACATAAACTGGAAAAGCCTCAGGAGCAACA	45100
Qy	201	ValSerGluGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys	220
Db	45099	GTCAAGCGAAGGGGGTTGTTCCCTGCTACTCTCCCATGTCACTTCAACCCCTTCTCTCAGAAG	45040
Qy	221	ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu	240
Db	45039	CCCCCTTACTCTTTCATTGACTGCAGTGTGAATATGGCTTGAACCTTTTCCCTTCATTA	44980
Qy	241	IleThrPheCysAsnSerSerLeuProValSerTrpArg	254
Db	44979	ATAACGTTTTGCAATTCATTCATTGCTGCCTGTCTCGTGAGG	44938

## RESULT 6

AX647177/C

## LOCUS

## DEFINITION

**ACCESSION**

VERSION  
KEYWORDS  
SOURCE  
ORGANIZATION

REFERENCE	AUTHOR	TITLE	JOURNAL
1	...	...	...
2	...	...	...
3	...	...	...
4	...	...	...
5	...	...	...
6	...	...	...
7	...	...	...
8	...	...	...
9	...	...	...
10	...	...	...
11	...	...	...
12	...	...	...
13	...	...	...
14	...	...	...
15	...	...	...
16	...	...	...
17	...	...	...
18	...	...	...
19	...	...	...
20	...	...	...
21	...	...	...
22	...	...	...
23	...	...	...
24	...	...	...
25	...	...	...
26	...	...	...
27	...	...	...
28	...	...	...
29	...	...	...
30	...	...	...
31	...	...	...
32	...	...	...
33	...	...	...
34	...	...	...
35	...	...	...
36	...	...	...
37	...	...	...
38	...	...	...
39	...	...	...
40	...	...	...
41	...	...	...
42	...	...	...
43	...	...	...
44	...	...	...
45	...	...	...
46	...	...	...
47	...	...	...
48	...	...	...
49	...	...	...
50	...	...	...
51	...	...	...
52	...	...	...
53	...	...	...
54	...	...	...
55	...	...	...
56	...	...	...
57	...	...	...
58	...	...	...
59	...	...	...
60	...	...	...
61	...	...	...
62	...	...	...
63	...	...	...
64	...	...	...
65	...	...	...
66	...	...	...
67	...	...	...
68	...	...	...
69	...	...	...
70	...	...	...
71	...	...	...
72	...	...	...
73	...	...	...
74	...	...	...
75	...	...	...
76	...	...	...
77	...	...	...
78	...	...	...
79	...	...	...
80	...	...	...
81	...	...	...
82	...	...	...
83	...	...	...
84	...	...	...
85	...	...	...
86	...	...	...
87	...	...	...
88	...	...	...
89	...	...	...
90	...	...	...
91	...	...	...
92	...	...	...
93	...	...	...
94	...	...	...
95	...	...	...
96	...	...	...
97	...	...	...
98	...	...	...
99	...	...	...
100	...	...	...

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

CDS
join(201..*246,25640..25677,27078..27094,141192..141769,
159571..159606,174525..174575,234891..235013,
235514..235560,279677..>279729)
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD70100.1"
/db_xref="GI:28800411"
/translation="MAASISASDAMDPTGCDVNAELPAAVATILGALLAAATLAVS
PLLVTLNRQRQEPHYLLPANILLSDLAYILLHMLISSSLGGWELGRWACGLIT
DAVFAACTSILSFATVLYHLYLAVIHPRLYLSMHSAAKAVALIWLVACCFPTFL
IWLKWDQAOLEQCAVILPMSMGTQPGCLLVITYTSLICVLCTALTJANGCFWR
LYAEAKTSVAGETQLCLHERNDPSDGGRRSKAFPLPGOLSSSCFTFRPOLRVLAHL
PIQGNDFNVSFSLGCLCPRLCPRLPVPELSPQAAASPCAMENIQRV"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="old seq 1369: original length 744.802 splitted into
3 new seqs--new seq 1369: from 000.001 to 349.989 *-new seq
2072: from 300.010 to 649.990 *-new seq 2073: from 600.011
to 744.802 *"

```

## ORIGIN

Alignment Scores:	
Pred. No.:	4, 11e-114
Score:	1299.00
Percent Similarity:	99.61%
Best Local Similarity:	99.61%
Query Match:	99.62%
DB:	6
Length:	349989
Matches:	253
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

UUS-09-441-857-12 (1-254) x AX647177 (1-349989)

Qy	1	MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly	20
Db	345708	ATGTCGCGGCGGAGACTTCGGAAATCCGCTGAGAAATTCAGCTGCTGTTCTCTGGGG	345649
Qy	21	GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn	40
Db	345648	GAGCAAAAGCGTTGCAAAAGACATCTTTTGATCATCCAGATTCAGGTATGACAGTTTTCACAAC	345899
Qy	41	ThrTyrGlnAlaIleIleGlyLeaAspPheLeuSerLysThrMetTyrLeuGluAspGly	60
Db	345588	ACCTATCAGGCAATAATTTGGCATTTGACTTTTTATCAAAAACCTATGCTACTTGGAGGATGGA	345529
Qy	61	ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro	80
Db	345528	ACAATCGGGCTTCGGCTGTGGGATACGGCGGGTCAGGAACGGTCTCCGTAGCCCTCATPTCC	345456
Qy	81	ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer	100
Db	345468	AGGTACATCCGTGATTTCTGCTGCGAGCTGTAGTAGTTTACGATATCACAAATGTTAACTCA	345409
Qy	101	PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle	120
Db	345408	TTTCAGCAAACTACAAAGTGGATTGATGATCTCAGAACAGAAAAGAGGAAGTGATGTTATC	345349
Qy	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu	140

## RESULT 6

AX647177/c

LOCUS	AX647177	349989 bp	DNA	Linear
1	1	1	1	1

DEFINITION Sequence 1369 from Patent EP1270724.

ACCESSION AX647177



```

Db      345348 ATCACCTAGTAGGAATAAGAACAGATCTTGTCACAAGGCAAGTGTGAGTTGAGGAG 345289
QY      141 GYGLuArgLyseAlaLyseGlyLeuAenValThrPheIleGluThrArgAlaLyseThrGly 160
Db      345288 GGNAGAGGAGNAGCCAAAGGGCTGAATGTACGTTATTGAACTAGGCGAAAGCTGGA 345229
QY      161 TyrAsnValLyseGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180
Db      345228 TACAATGTAAAGCAGCTCTTTTCAGCTGTAGCAGCAGCTTTGCCGGAATGGAAGCACA 345169
QY      181 GlnAspGlySerArgGluAspMetSerAspIleLyseLeuGluLyseProGlnGluGlnThr 200
Db      345168 CAGGACGGAGCAGAGAAACATGAGTGACATAAAACTGGAAGAGCTCAGGAGCAACA 345109
QY      201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerThrLeuProGlnLyse 220
Db      345108 GTCAGGAGAGGGGGTGTTCCTGCTACTCTCCATGTCATCTTCAACCCCTTCTCAGAG 345049
QY      221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240
Db      345048 CCCCCTTACTCTTTTCACTGACGTCAGTGTGAATATTGGCTTGAACCTTTTCCCTTCATTA 344989
QY      241 IleThrPheCysAsnSerSerLeuLeuProValSerTyrArg 254
Db      344988 ATAACGTTTGGCAATTCAATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 344947

RESULT 7
AC018865      165411 bp      DNA      linear      PRI 23-NOV-2002
LOCUS      Homo sapiens BAC clone RP11-32C20 from 2, complete sequence.
AC018865
ACCESSION      AC018865
VERSION      AC018865.4
KEYWORDS      GI:25189068
SOURCE      HTG.
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 165411)
AUTHORS      Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 165411)
AUTHORS      Kim, K., Haglund, K., Tomlinson, C., Mangiapanello, L. and Dignan, G.
TITLE      The sequence of Homo sapiens BAC clone RP11-32C20
JOURNAL      Unpublished (2001)
REFERENCE      3 (bases 1 to 165411)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (21-DEC-1999) Genome Sequencing Center, Washington
REFERENCE      4 (bases 1 to 165411)
AUTHORS      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
TITLE      MO 63108, USA
JOURNAL      Submitted (13-SEP-2002) Genome Sequencing Center, Washington
REFERENCE      5 (bases 1 to 165411)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (23-NOV-2002) Department of Genetics, Washington
REFERENCE      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS      On Nov 23, 2002 this sequence version replaced gi:22830544.
TITLE      ----- Genome Center
JOURNAL      Center: Washington University Genome Sequencing Center
COMMENT      Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0032C20
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

There is an unresolved tandem repeat from base 115267 to 124390.

Polymorphisms have been identified between AC018804 and this sequence.

Data from AC018804 was used to finish this clone.

This sequence is the entire insert of the clone.

FEATURES	source
Location/Qualifiers	1..165411
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-32C20"
	/clone_lib="RPCI-11"
	487..1684
repeat_region	/rpt_family="L1"
	1678..1885
repeat_region	/rpt_family="L1"
	1890..1986
repeat_region	/rpt_family="L1"
	1987..2289
repeat_region	/rpt_family="Alu"
	2294..2380
repeat_region	/rpt_family="GA-rich"
	2392..2699
repeat_region	/rpt_family="L1"
	2905..3262
repeat_region	/rpt_family="Alu"
	3284..3395
repeat_region	/rpt_family="T-rich"
	3601..3674
repeat_region	/rpt_family="MER105"
	3726..3845
repeat_region	/rpt_family="Alu"
	3846..3869
repeat_region	/rpt_family="(TAA)n"
	3870..4042
repeat_region	/rpt_family="Alu"
	4064..4364

```
repeat_region /rpt_family="ERV1"
4692..4980
repeat_region /rpt_family="Alu"
4981..5001
repeat_region /rpt_family="AT_rich"
5006..5316
repeat_region /rpt_family="Alu"
5340..5519
repeat_region /rpt_family="ERV1"
5563..5604
repeat_region /rpt_family="AT_rich"
5637..5678
repeat_region /rpt_family="AT_rich"
5757..5781
repeat_region /rpt_family="AT_rich"
5782..6005
repeat_region /rpt_family="Alu"
6067..6146
repeat_region /rpt_family="MER1_type"
6196..6487
repeat_region /rpt_family="Alu"
6515..6634
repeat_region /rpt_family="MER1_type"
6755..7044
repeat_region /rpt_family="Alu"
7349..7641
repeat_region /rpt_family="Alu"
7644..7932
repeat_region /rpt_family="Alu"
7948..8108
repeat_region /rpt_family="BC200"
8294..8589
repeat_region /rpt_family="Alu"
8824..9036
repeat_region /rpt_family="L1"
9037..9342
repeat_region /rpt_family="Alu"
9343..9404
repeat_region /rpt_family="L1"
9405..9444
repeat_region /rpt_family="Alu"
9445..9472
repeat_region /rpt_family="L1"
9473..9748
repeat_region /rpt_family="Alu"
9749..9903
repeat_region /rpt_family="L1"
10276..10381
repeat_region /rpt_family="MIR"
10488..10516
repeat_region /rpt_family="L1"
10517..10806
repeat_region /rpt_family="Alu"
10918..10942
repeat_region /rpt_family="L1"
11350..11603
repeat_region /rpt_family="Alu"
11614..11738
repeat_region /rpt_family="MIR"
11961..12196
repeat_region /rpt_family="MIR"
12222..12355
repeat_region /rpt_family="CR1"
12365..12448
repeat_region /rpt_family="MIR"
12449..12747
repeat_region /rpt_family="Alu"
12802..12853
repeat_region /rpt_family="L1"
12855..13172
repeat_region /rpt_family="Alu"
13199..13700
repeat_region /rpt_family="L1"

repeat_region 14431..14726
/rpt_family="Alu"
14754..14809
/rpt_family="MIR"
15272..15400
/rpt_family="CT-rich"
15425..15458
/rpt_family="(TTCTC)n"
15465..15650

Alignment Scores:
Pred. No.: 5.91e-114 Length: 165411
Score: 1293.00 Matches: 252
Percent Similarity: 99.21% Conservative: 0
Best Local Similarity: 99.21% Mismatches: 2
Query Match: 99.16% Indels: 0
DB: 9 Gaps: 0

US-09-441-857-12 (1-254) x AC018865 (1-165411)
Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 36142 ATGTCCGGCGCGAGACTTCGGGAATCCGCTCAGGAAATTCAGAGCTGGTCTCTCGGG 36201
Qy 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40
Db 36202 GAGCAAGCGTTCGAAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTGTGCAAC 36261
Qy 41 ThrTyrGlnAlaLeuLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 36262 ACCTATCAGGCAATAATTGGCATTGACTTTTATCAAAAATATGATCTTTGGAGATGGA 36321
Qy 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuLeuPro 80
Db 36322 ACAATCGGGTTCGGCTGTCGGATACGGCGGTCAGGAACGCTCTCCGTAGCTCATTCCTCC 36381
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValValTyrAspIleThrAsnValAsnSer 100
Db 36382 AGGTACATCCGTGATTCTGCTCAGCTGTAGTAGTTTACCATATCAGAAATGTTTAACTCA 36441
Qy 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 36442 TTCAGCAAACTACAAAGTCGATTGATGATGTAGACACACAGAGGCAAGTGTGATGTATC 36501
Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 36502 ATCAGCTAGTAGGAAATAGAACAGATCTTGTCTCACAAGAGGCAAGTGTGATGTAGGAG 36561
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 36562 GGAGAGAGGAAAGCCAAAGGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAAGCTGGA 36621
Qy 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180
Db 36622 TACAATGTAAGCAGCTCTTTTCGACGCTGAGCAGCAGCTTTCGGGGAATGGAAGCACA 36681
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 36682 CAGGACGGAAGCAGAGAACATGAGTGACATAAACTGGAAGAGCCCTCAGGAGCAAAACA 36741
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
Db 36742 GTCAGCGAAGGGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 36801
Qy 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240
Db 36802 CCCCCCTACTCTTTTCATTGATCGAGTGAGTAATTTGGCTTGAACCTTTTCCCTTCATTA 36861
Qy 241 IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254
Db 36862 ATAACGTTTTGCAATTCATCATTTGCTGCCTGTCTCGTGGAGG 36903

RESULT 8
```

AC073869/c  
 LOCUS AC073869 195280 bp DNA linear PRI 09-JAN-2002  
 DEFINITION Homo sapiens BAC clone RP11-109E12 from 2, complete sequence.  
 ACCESSION AC073869  
 VERSION AC073869.5 GI:15638756  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Suleston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074

REFERENCE  
 AUTHORS Bielicki,L., Maupin,R. and Meyer,R.  
 TITLE The sequence of Homo sapiens BAC clone RP11-109E12  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 195280)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Sep 18, 2001 this sequence version replaced gi:14209794.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0109E12  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong

and coworkers at the Roswell Park Cancer Institute  
 (http://bacpac.med.buffalo.edu)  
 VECTOR: pBACe3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the right is RP11-433A19. Actual start of this clone is at base position 1 of RP11-109E12; actual end is at base position 195280 of RP11-109E12.

Data from AC022629 was used to finish AC073869.

FEATURES  
 Location/Qualifiers  
 1..195280  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-109E12"  
 /clone\_lib="RPCI-11"  
 1..167  
 /note="match to EST AA904412 (NID:g3039535) ok07e01.s1"  
 82..96  
 /note="similar to Homo sapiens EST AV651649 (NID:g9872663)"  
 298..778  
 /note="match to EST AW293618 (NID:g6700254)"  
 408..431  
 /note="similar to Homo sapiens EST AV651649 (NID:g9872663)"  
 503..670  
 /note="match to EST AI352084 (NID:g4089290) qr10b01.x1"  
 651..1100  
 /note="match to EST AI138580 (NID:g3644552) qd85b12.x1"  
 695..710  
 /note="similar to Homo sapiens EST BE791748 (NID:gl0212946)"  
 1698..2014  
 /note="similar to Homo sapiens EST N30408 (NID:gl148928) YV13C09.s1"  
 2015..2410  
 /note="similar to Homo sapiens EST AV651649 (NID:g9872663)"  
 2253..2732  
 /note="match to EST AW340014 (NID:g6836640) hc91e04.x1"  
 2338..3119  
 /note="similar to Homo sapiens EST AUL139340 (NID:gl1000861)"  
 2340..2885  
 /note="similar to Bos taurus EST BE722624 (NID:gl0123920)"  
 3154..3941  
 /note="similar to Canis familiaris EST B1430421 (NID:gl15207653)"  
 3236..3645  
 /note="similar to Mus musculus EST AA617192 (NID:g2504397) VJ79a09.r1"  
 4375..4376  
 /note="similar to Homo sapiens EST AI677857 (NID:g4888039) wd34c04.x1"  
 4427..4468  
 /rpt\_family="(TTTA)n"  
 4440..4738  
 /rpt\_family="Alu"  
 4835..4973  
 /note="similar to Homo sapiens EST AI677857 (NID:g4888039) wd34c04.x1"  
 4892..5127  
 /note="similar to Homo sapiens EST BF511460 (NID:gl1594758)"  
 4899..5106  
 /note="match to EST AW804021 (NID:g7855891)"  
 4974..5130  
 /note="similar to Homo sapiens EST AI677857 (NID:g4888039) wd34c04.x1"  
 5205..5387  
 /note="similar to Homo sapiens EST AI677857 (NID:g4888039) wd34c04.x1"







```

/protein_id="AAH03617.1"
/db_xref="GI:13177664"
/db_xref="L:13177664"
/db_xref="MM:129513"
/tax_id="12846"
/translation="MSTGDFGNPLKPLVFLPGSVGKTSLTIRPMVDSFNTYCA
TIGDLSKTMVLIEDTIRLQWDATAGERFSLIPSVIRDSAAVYVYDITNVN9G
QTKKQDDVTRGERSVIMLVGNKTDLADKRGSLIGERKGAELNMIETSAKAG
NVKPLGRVAALPMSESTPDSREDMTDIDLEKVEPQSPFGKNCVF"

```

## ORIGIN

```

Alignment Scores:
  Pred. No.:      5.9e-102      Length:      3070
  Score:          1147.00      Matches:      228
  Percent Similarity: 90.70%    Conservative:  6
  Best Local Similarity: 88.37%  Mismatches:   20
  Query Match:      87.96%     Indels:       4
  DB:                9          Gaps:         1

US-09-441-857-12 (1-254) x BC0031617 (1-3070)

```



CDS	/db_xref="MIM:179513"		/db_xref="MIM:179513"		ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236	
	463..1089		463..1089			
	/gene="RAB6A"		/gene="RAB6A"			
	/codon_start=1		/codon_start=1			
misc_feature	/product="RAB6A protein"		/product="RAB6A protein"			
	/protein_id="AAH68486.1"		/protein_id="AAH68486.1"			
	/db_xref="GI:46249771"		/db_xref="GI:46249771"			
	/db_xref="LocusID:5870"		/db_xref="LocusID:5870"			
ORIGIN	/db_xref="MIM:179513"		/db_xref="MIM:179513"			
	/translation="MSTGGDFGNPLRKFLVFLGQSVGKTSLITRFMYDSFDNTYQA		/translation="MSTGGDFGNPLRKFLVFLGQSVGKTSLITRFMYDSFDNTYQA			
	TIGIDFLSKTYLEDTRLQLMDTAGQERFRSLIPYIRDSAAAAYVYDITNVNSFQ		TIGIDFLSKTYLEDTRLQLMDTAGQERFRSLIPYIRDSAAAAYVYDITNVNSFQ			
	YNNKQIDVVRTERGSDVIMLVGNKTKDLADKROVSEBGRKAKELNWMFTETSAKAG		YNNKQIDVVRTERGSDVIMLVGNKTKDLADKROVSEBGRKAKELNWMFTETSAKAG			
Alignment Scores:	502..987		502..987			
	/gene="RAB6A"		/gene="RAB6A"			
	/note="RAB; Region: Rab subfamily of small GTPases"		/note="RAB; Region: Rab subfamily of small GTPases"			
	/db_xref="CDD:cd00154"		/db_xref="CDD:cd00154"			
Pred. No.:	5.98e-102		5.98e-102			
	Length: 3100		Length: 3100			
	Matches: 228		Matches: 228			
	Conservative: 6		Conservative: 6			
Percent Similarity:	90.70%		90.70%			
	Best Local Similarity: 88.37%		Best Local Similarity: 88.37%			
	Mismatches: 20		Mismatches: 20			
	Indels: 4		Indels: 4			
Query Match:	87.96%		87.96%			
	DB: 9		DB: 9			
	Gaps: 1		Gaps: 1			
	Gaps: 1		Gaps: 1			
US-09-441-857-12 (1-254) x BC068486 (1-3100)						
Qy	1	MetSerAlaGlyIleGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20				
Db	463	ATGTCACGGGCGGAGACTTCGGGAAATCGCTGAGGAAATTCAGCTGGTTCCTGGGG 522				
Qy	21	GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40				
Db	523	GAGCAAAAGCGTTGGAAGACATCTTTGATCACCAGATTCAATGATGACAGTTTGGACAA 582				
Qy	41	ThrTyrGlnAlaIleIleGlyIleAspPheLysSerLysThrMetTyrLeuGluAspGly 60				
Db	583	ACCTATCAGGCAACAAATTTGGCATTTGATCAAAATATGATCTTCGGAGATCGA 642				
Qy	61	ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80				
Db	643	ACAAATCAGCTTCAGCTCGGATCTCGGGTCAAGAACGTTTCCGTAGCTCATTTCCC 702				
Qy	81	ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer 100				
Db	703	AGTTACATCCGTGATTCTGCTGCAGCTGTAGTAGTTACGATATCACAAATGTTAACTCA 762				
Qy	101	PheGlnGlnThrThrLysThrIleAspAspValArgThrGluArgGlySerAspValIle 120				
Db	763	TTCCAGCAAACTACAAAGTGGATTGATGTCTGACACAGAAAGGAGGATGTATTATC 822				
Qy	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140				
Db	823	ATCATGCTAGTAGGAAATAAAACAGATCTTCTGACACAGAGCAAGTGTCAATTGAGGAG 882				
Qy	141	GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160				
Db	883	GGAGAGAGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTGCACAAAGCTGGA 942				
Qy	161	TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180				
Db	943	TACAAATGAAAGCAGCTCTTTCGACGCTGTACAGCAGCTTTGCCGGAAGTGAAGACCA 1002				
Qy	181	GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200				
Db	1003	CAGGACAGAAAGCAGAGAGATATGATTGACATATAAACTGGAAAGCTTCAGGACCAACCA 1062				
Qy	201	ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerThrLeuProGlnLys 220				
Db	1063	GTCAGTGAAGAGGCTGTTCCTGCTAATCTCCCATGTCTCAACCTTCTTCAGAAAGC 1122				

Qy	221	-----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236			
Db	1123	TCACTGCTTGGCCCCCTTACTCTTTTCATTGACTGCAGTGTGAATATTGCTTGAACCTT 1182			
Qy	237	PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254			
Db	1183	TTCCCTTCAGTAATAACGATTATGCAATTCATTCGTGCTGCTCTCGTGGAGA 1236			
RESULT 13					
AK057157					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Hirakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Foglia, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Toghiani, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.	Complete sequencing and characterization of 21,243 full-length human cDNAs				
JOURNAL					
PUBMED					
REFERENCE					
AUTHORS					
Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.	NEDO human cDNA sequencing project				
Unpublished					
3 (bases 1 to 3120)					
Isogai, T., Otsuki, T. and Sugiyama, T.					
Direct Submission					
Submitted (24-OCT-2001)					
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan					
(E-mail:genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)					
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library					



Qy	1	MetSerAlaGlyGlyAspPheGlyAsnProLeuArgGlyPheIysLeuValPheLeuGly	20
Db	487	ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGACTGTGTTCCTCGGGG	546
Qy	21	GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn	40
Db	547	GAGCAAAAGGTTGGAAAAGACATCTTTGATACCCAGATTTCATGTATGACAGTTCCTTGACAAC	606
Qy	41	ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerIysThrMetTyrLeuGluAspGly	60
Db	607	ACCTATCAGCCAAACATTTGGCATTCATCTTTATCAAAACTATGTAATCTTGAGGATCGA	666
Qy	61	ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro	80
Db	667	ACAAATCAGGCTTCAGCTGTGGGATCTGCGGGTCAGAACGTTTCGTAGCCTCATCC	726
Qy	81	ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer	100
Db	727	AGTTACATCCGTGATTCTGCTGCAGCTGTAGTAGTTTACGATATCACAAATGTTAACTCA	786
Qy	101	PheGlnGlnThrThrLysTrpIleAspValArgThrGluArgGlySerAspValIle	120
Db	787	TTCCAGCAAACTACAAAGTGGATTGATGATGTCTAGAAACAGAGAGGAAAGTGATGTATC	846
Qy	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspIysArgGlnValSerValGluGlu	140
Db	847	ATCATGCTAGTAGGAAATTAACAACAGATCTCTCTCAAGAGGCAAGTGTCAAATTGAGGAG	906
Qy	141	GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly	160
Db	907	GGAGAGAGGAAAGCCAAAGACTGAATGTTATGTTTAAACTAGTGCAAAAGCTGGA	966
Qy	161	TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr	180
Db	967	TACAATGTAAACAGACTCTTTTCAGCTGTACAGCAGCTTTGCGGGATGGAAACACA	102
Qy	181	GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluInThr	200



```
Db 391 ACCTATCAGCGCAACAATTGGCATTGACCTTTATCAAAAACAATGTACTTGGAGGATAGA 450
QY 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 451 ACCATCAGGCTCAGCTGTGGGATACTCGGGTCAGGAACGTTTCGTAGCCTCATTTCCC 510
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 511 AGTTACATCCGTGACTCTGCTCAGCGGTAGTAGTTTACGATATCACAAATGTTAACTCA 570
QY 101 PheGlnGlnThrThrIysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 571 TTCCAGCAAACTACAAAATGGAATTGATGATGTGCAGAACAGAGAGAGGAGTGTGTCATC 630
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 631 ATCATGCTAGTAGGAATAAACAAGATCTTGTGTATAGAGGCAAGTGTCAATTGAGGAG 690
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 691 GGAGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTTATTTGAAACCAGTGC AAAAGCAGGA 750
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGlnSerThr 180
Db 751 TACAATGTAAAGCAGCTTTCCGGCGTGTTCGTCAGCTTTACCTGGAATGGAAGCACA 810
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200
Db 811 CAGGACAGAAAGCAGAGAACATGATTGACATAAAACTGGAAGAGCCTCAGGAGCAACCA 870
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
Db 871 GTCATAGAGCGCGCTGTTCTCTGCTAATCCCTCGCATCTTTCAACCCCTCTGCAGAG 930
QY 221 -----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236
Db 931 CTCACCTGCTTTGGCCCCCATACTCTTTCAATTGACTGCAGTGTGAATATTGGCTTGAACCTT 990
QY 237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254
Db 991 TTCCCTTCAGTAATAACGATTGCAATTTCATTCCTGCCTGTTTCGTGGAGG 1044
```

Search completed: April 25, 2005, 04:55:04  
Job time : 3685 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2005, 22:09:40 ; Search time 452 Seconds  
(without alignments)  
3326.580 Million cell updates/sec

Title: US-09-441-857-12

Perfect score: 1304

Sequence: 1 MSAGDGFNPLRKLKFLVFLG.....NLFPSLITFCNSSLIPVSWR 254

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues.

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-Q=/Cgn2\_1/USPTO\_spool\_p/US0441857/runat\_22042005\_155508\_24737/app\_query.fasta\_1.391  
-DB=N Geneseq\_16Dec04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MOD=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09441857@CN\_1\_1\_708@runat\_22042005\_155508\_24737 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq\_16Dec04:.\*  
1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002as:.\*  
7: Geneseqn2002bs:.\*  
8: Geneseqn2003as:.\*  
9: Geneseqn2003bs:.\*  
10: Geneseqn2003cs:.\*  
11: Geneseqn2003ds:.\*  
12: Geneseqn2004as:.\*  
13: Geneseqn2004bs:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1304	100.0	762	AA293839	AA293839 WTH3 codi
2	1304	100.0	3084	ABX71303	ABX71303 Human int
C 3	1299	99.6	349981	ADC87619	ADC87619 Human GPC
C 4	1299	99.6	349989	ADC86916	ADC86916 Human GPC
5	1147	88.0	1460	AA293840	AA293840 RAB6C cod

6	1147	88.0	3079	12	ADN03693	Adn03693 Antipori
7	1143	87.7	719	3	AAZ93834	Aaz93834 WTH3 5' p
8	1138	87.3	3745	12	ADL12658	Adl12658 Human ste
9	1133	86.9	3118	3	AAF15989	Aaf15989 Human pro
10	982	75.3	739	3	AAZ93836	Aaz93836 RAB6C cod
11	959	73.5	718	3	AAZ93835	Aaz93835 RAB6 codi
12	959	73.5	740	3	AAZ93837	Aaz93837 RAB6 codi
13	953	73.1	624	3	AAZ93841	Aaz93841 RAB6C cod
14	892	68.4	1219	4	AAK52711	Aak52711 Human pol
15	892	68.4	1266	12	ADP07320	Adp07320 Human RAB
16	892	68.4	1266	13	ADR25795	Adr25795 Breast ca
17	892	68.4	1461	4	AAK51727	Aak51727 Human pol
18	884	67.8	1589	2	AAZ11736	Aaz11736 Human tra
19	883	67.7	573	3	AAZ93838	Aaz93838 WTH3 codi
20	865.5	66.4	3195	5	AA887692	Aa887692 DNA encod
21	846.5	64.9	1188	4	ABL11393	Ab111393 Drosophil
22	846.5	64.9	1188	4	ABL11392	Ab111392 Drosophil
23	835.5	64.1	2456	10	ADA52862	Ada52862 Human cod
24	715	54.8	926	3	AAF14485	Aaf14485 Aspergill
25	712	54.6	1074	3	AAC33337	Aac33337 Arabidops
26	709.5	54.4	1355	3	AAC46575	Aac46575 Zea mays
27	673.5	51.6	624	3	AAC42719	Aac42719 Arabidops
28	667	51.2	523	13	ACN59588	Acn59588 Cotton gy
29	665	51.0	636	10	ADC76216	Adc76216 DNA homol
30	665	51.0	636	10	ADK55884	Adk55884 Plant DNA
31	643	49.3	645	3	AAC43018	Aac43018 Arabidops
32	642	49.2	443	8	ABX42190	Abx42190 Bovine ES
33	627	48.1	646	10	ADD16517	Add16517 DNA (Seqi
34	627	48.1	646	11	ADM44783	Adm44783 Insect re
35	627	48.1	647	10	ADC75668	Adc75668 DNA homol
36	627	48.1	723	10	ADC76194	Adc76194 DNA homol
37	627	48.1	723	10	ADK58490	Adk58490 Plant DNA
38	612.5	47.0	495	3	AAC41466	Aac41466 Zea mays
39	596	45.7	791	13	ADR65699	Adr65699 Cotton cd
40	596	45.7	928	6	ABI99630	Abi99630 Mouse ioc
41	593.5	45.5	3826	10	ADC86930	Adc86930 Human GPC
42	586.5	45.0	642	5	AA887691	Aa887691 DNA encod
43	582	44.6	648	13	ADS47056	Ads47056 Bacterial
44	572	43.9	846	13	ADS47382	Ads47382 Bacterial
45	569	43.6	424	9	ACH18328	Ach18328 Human adu

ALIGNMENTS

RESULT 1  
AAZ93839  
ID AAZ93839 standard; DNA; 762 BP.

AC AAZ93839;

DT 29-AUG-2000 (first entry)

DE WTH3 coding sequence.

XX WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
KW antibody; immunogen; mutation; detection; therapy; human; ss.  
XX Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 1..762  
FT /\*tag= a  
FT /product= "WTH3"

XX PN WO200029625-A1.

XX PD 25-MAY-2000.

XX PF 18-NOV-1999; 99WO-US027630.

XX PR 18-NOV-1998; 98US-0108994P.

XX PA (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;  
XX WPI; 2000-387828/33.  
DR P-PSDB; AAY83649.  
XX Novel nucleic acids encoding wh3 protein useful for inhibiting multiple  
PT drug resistance in tumor cells for identifying the modulators of drug  
PT resistance.  
XX Claim 3; Page 62-64; 82pp; English.  
XX Nucleic acids encoding WTH3 and RAB6 were isolated by Methy1- (or mutant)  
CC Differential Display (MDD) methods and can be used in methods for  
CC detecting methylation patterns in multiple drug resistance loci. Genes  
CC are frequently not methylated in cells where they are expressed but are  
CC methylated in cell types where they are not expressed. Tumour cell DNA is  
CC often methylated to a different extent and in different regions when  
CC compared to DNA of normal cells. The methylation pattern in a multiple  
CC drug resistance locus can be altered and give rise to altered expression  
CC patterns of that multiple drug resistance locus. Nucleic acids  
CC corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
CC can be used as probes for detecting mutations and methylation patterns of  
CC those loci. The nucleic acids and their homologues are useful for  
CC inhibition of multiple drug resistance and for treating tumors exhibiting  
CC multiple drug resistance. They are also useful for detecting and  
CC measuring the expression of mRNA from identified genes and for  
CC determining suitable therapeutic treatment. Antibodies directed against  
CC immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
CC proteins and polypeptides in tissues or body fluids of patients  
XX Sequence 762 BP; 221 A; 157 C; 194 G; 190 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,01e-139 Length: 762  
Score: 1304.00 Matches: 254  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-441-857-12 (1-254) x AAZ93839 (1-762)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 1 ATGTCGCGCGCGGAGACTTCGGGAATCCGTCAGGAAATTCAGACTGGTCTCTCTGGGG 60  
Qy 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40  
Db 61 GAGCAAGCGTGGNAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTCACAC 120  
Qy 41 ThrTyrGlnAlaIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 121 ACCTATCAGGCAATAATGGCATTGACATTTTATCAAAATATGATCTTCGAGAGTGA 180  
Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
Db 181 ACAATCGGGCTTCGGCTGTGGGATACCGCGGGTCAGGAACGTCTCCGTAGCTCATTTCCC 240  
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
Db 241 AGGTACATCCGTGATCTGCTGAGCTGTAGTATTACGATATACAAATGTTAACTCA 300  
Qy 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120  
Db 301 TTCCAGCAAACTACAAAGTGGATTGATGATCTCAGAACAGAAAGAGGAGTGTATTATC 360  
Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
Db 361 ATCAGCTAGTAGGAAATAGAACAGATCTCTCTCAGAGGCAAGTGTCTAGTTGAGGAG 420  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160

Db 421 GGAGAGAGAAAGCCARAGGGCTGAATGTTTACGTTTATTGAAACTAGGCGCAAAACTGGA 480  
Qy 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
Db 481 TACAATGTAAGCAGCTCTTTTCGACGCTGTAGCAGCAGCTTTCGCGGGAATGAAAGCACA 540  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
Db 541 CAGACCGAGAGCAGAGAGACATGAGTGACATAAACTGGAAAAGCCTCAGGAGCAACA 600  
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220  
Db 601 GTCAGCGAAGGGGTGTTCTCTGCTACTCTCCCATGTCACTTCAACCTTCTCTCAGAG 660  
Qy 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240  
Db 661 CCCCTTACTCTTTTCAATGACTGCAGTGTGAATATTCGCTTGAACCTTTTCCCTTCATTA 720  
Qy 241 IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254  
Db 721 ATAACGTTTTGCAATTCATTCGCTGCTCTCGTGGAGA 762

RESULT 2

ID ABX71303 standard; cDNA; 3084 BP.

XX AC ABX71303;

XX DT 14-APR-2003 (first entry)

XX DE Human intracellular transport and trafficking cDNA DKFZphkd2\_4k14.

XX KW Human; gene; gene therapy; vaccine; disease treatment; detection; ss.

XX OS Homo sapiens.

XX PN W0200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.

XX PR 28-SEP-1999; 99US-0156503P.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX PI Wiemann S;

XX DR WPI; 2001-327840/34.

XX DR P-PSDB; ABUS2844.

XX PT Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

XX PS Claim 1; Page 432-433; 1095pp; English.

XX CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention

SQ Sequence 3084 BP; 881 A; 665 C; 662 G; 876 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1.53e-138 Length: 3084  
Score: 1304.00 Matches: 254



Db	45279	GGAGAGGAAAGCCAAAGGCGTAAATGTTACGTTTATTGAAACTAGGGCAAAAGCTCGA	45220
Qy	161	TyrAsnValIysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr	180
Db	45219	TACAATGTAAGACAGCTCTTTTCAGCTGTAGCAGCAGCTTTTCGCGGAATGGAAGCACA	45160
Qy	181	GlnAspClySerArgGluAspMetSerAspLeIeLysLeuGluIysProGlnGluGlnThr	200
Db	45159	CAGCAGGAAGCAGAGAAGACATGAGTGACATAAACTCGAAAGCCTCAGAGCAACA	45100
Qy	201	ValSerGluGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnIys	220
Db	45099	GTCAAGCAAGGCGGTGTTCCTGCTACTTCCCATGTCTTCAACCCCTCCTCAGAAG	45040
Qy	221	ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu	240
Db	45039	CCCCCTACTCTCTTTCATCTGACTCGAGTGTGAATATTTGGCTTTTGAACTTTTCCCTTCATTA	44980
Qy	241	IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg	254
Db	44979	ATAACGGTTTTGCAATTTCATTCATGCTGCCTGTCTCGTGAGG	44938

#### RESULT 4

ADC86916/c  
ID ADC86916 standard; DNA; 349989 BP.

01-JAN-2004 (first entry)

DE Human GPCR gene SEO ID NO:1369.

KW ds; gene; human; GPCR;  
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.

KW guanosine triphosphate-binding protein coupled receptor; gene therapy.

OS Homo sapiens.

AX  
PN  
EP1270724-A2.

PD 02-JAN-2003.

18-JUN-2002; 2002EP-00013517.

PR 18-JUN-2001: 2001JP-00246789.

PA (NAAD-) NAT INST ADVANCED IND SCI &amp; TECHNOLOGY.

PA (ADSC-) CENT ADVANCED SCI &amp; TECHNOLOGY INCUBATIO.

PI Suwa M, Asai K, Akiyama Y, Aburatani H;

WPI; 2003-315783/31.

DR P-PSDB; ADC86917.

New polynucleotide, useful for preparing a composition for treating a patient in need of increased or suppressed activity or expression of the guanosine triphosphate-binding protein coupled receptor.

PS Claim 1; SEQ ID NO 1369; 28pp; English.

The invention relates to a novel polynucleotide encoding a guanosine triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of the invention may have a use in gene therapy. The polynucleotide and the polypeptide are useful for preparing a composition for treating a patient in need of increased or suppressed activity or expression of the guanosine triphosphate-binding protein coupled receptor. The polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the invention.

SQ Sequence 349989 BP; 93096 A; 80097 C; 83967 G; 92327 T; 0 U; 502 Other;

Alignment Scores:

Pred. No.:	5.51e-135	Length:	349989
------------	-----------	---------	--------

Score:	1299.00	Matches:	253
--------	---------	----------	-----

Percent Similarity:	99.61%	Conservative:	0
Best Local Similarity:	99.61%	Mismatches:	1
Query Match:	99.62%	Indels:	0
DB:	10	Gaps:	0
US-09-441-957-12 (1-254) x ADC86916 (1-349989)			
Qy	1	MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly	20
Db	345708	ATGTCCGGCGGCGAGACTTCGGGAATCCGCTGAGGAAATCAAGTGGTGTTCTCTGGG	345649
Qy	21	GlunSerValAlaLysThrSerLeulleThrArgPheArgTyrrAspSerPheAspAsn	40
Db	345648	GAGCAAACGCGTGCAGAACATCTTTGATCACCAAGATTCAAGTATGACAGTTTTTGACAAC	345589
Qy	41	ThrTyrrGlnAlaIlelleGlyVileAspPheLeuSerLysThrMetTyrrLeuGluAspGly	60
Db	345588	ACCATCAGGCANPAATGGCATTTGACTTTTATCAAAAATACTGTACTTGGAGGATGGA	345529
Qy	61	ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuillePro	80
Db	345528	ACAATCGGCTTCGGCTGTGGATACGGCGGTGAGGAACGTCCTCCGTAGCTCATTCCTCC	345469
Qy	81	ArgTyrrIleArgAspSerAlaAlaValValValTyrrAspIleThrAsnValAsnSer	100
Db	345468	AGGTACATCCGTGATCTCGCTGCAGCTGTAGTAGTTTACGATATCACAAATGTTTAACFCA	345409
Qy	101	PheGlnGlnThrThrLysTsrPileAspAspValArgThrGluArgGlySerAspValIle	120
Db	345408	TTCAGCAAACTACAAAGTGATGTAGTAGTGTACAGACAGANAAGAGAAGTGATGTATC	345349
Qy	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValcluglu	140
Db	345348	ATCACGCTAGTAGGAATAGAACAGATCTTGCTCACAGAGGCCAAGTGTCA GTTGAGGAG	345289
Qy	141	GlyGluArgLysAlaLysGlyLeuAsnValThrPhelleGluThrArgAlaLysThrGly	160
Db	345288	GGAGAGAGGAAAGCCAAAGGCTGAAATGTTACGTTTATTGAAACTAGGGCCAAAAAGCTGGA	345229
Qy	161	TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr	180
Db	345228	TACAAATGTAAAGCAGCTCTTCGACGTGTAGCAGACGCTTTGCCGGGAATGGAAGACACA	345169
Qy	181	GlnAspGlySerArgGluAspMetSerAspileLysLeuGluLysProGlnGlnThr	200
Db	345168	CAGACCGAAGCAGAGAAGACATGAGTGACATAAACTGGAAAAGCCTCAGAGACAAACA	345109
Qy	201	ValserGluGlyCysSerCysTyrrSerProMetSerSerSerThrLeuProGlnLys	220
Db	345108	GTCAGCGAAGGGGTTGCTGCTACTCTCCCATGTATCTTCAACCCCTTCCTCAGAAG	345049
Qy	221	ProProTyrrSerPhelleAspCysSerValAsnIleGlyLeuAenLeuPheProSerLeu	240
Db	345048	CCCCCTTACTCTTTTCATTGACTGCAGGTGAATATGGCTTGAACCTTTTCCCCTTCATTA	344989
Qy	241	IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg	254
Db	344988	ATAACGTTTGGCAATTCATTCATTGCTGCCTGTCTCGTGGAGG	344947
RESULT 5			
ID	AAZ93840		
ID	AAZ93840 standard; DNA; 1460 BP.		
XX	AAZ93840;		
AC	AAZ93840;		
XX			
DT	29-AUG-2000 (first entry)		
XX			
DE	RAB6C coding sequence.		
XX			
KW	WH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;		
KX	antibody; immunogen; mutation; detection; therapy; human; ss.		
XX	Homo sapiens.		





CC The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polynucleotides of the invention.

XX  
SQ Sequence 3079 BP; 901 A; 639 C; 636 G; 903 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1.49e-120 Length: 3079  
Score: 1147.00 Matches: 228  
Percent Similarity: 90.70% Conservative: 6  
Best Local Similarity: 88.37% Mismatches: 20  
Query Match: 87.96% Indels: 4  
DB: 12 Gaps: 1

US-09-441-857-12 (1-254) x ADN03693 (1-3079)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 427 ATGTCACGGCGGAGACTTCGGGAATCGCTGAGAAATTCAGCTGGTGTCTCTGGGG 486  
Qy 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40  
Db 487 GAGCAAGCGTTGGAAAGACATCTTTGATCACCAGATTTCATGTATGACAGTTTTGACAA 546  
Qy 41 ThrTyrGlnAlaIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 547 ACCATATCAGGCAACAATTTGGCATTGACTTTTATCAAAAACATATGTACTTGAGGATCGA 606  
Qy 61 ThrLeuGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuLeuPro 80  
Db 607 ACAATCAGCTTCAGCTGGGATCTCGGGTACGGAACGTTCCGTAGCTCTATTC 666  
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
Db 667 AGTTACATCCGCGATTCTGCTGCAGCTGTAGTTAGTTTACGATATCACAAATGTTAACTCA 726  
Qy 101 PheGlnGlnThrThrLysThrIleAspAspValArgThrGluArgGlySerAspValIle 120  
Db 727 TTCACGCAAACTTACAAAGTGGATGATGATGTCAGAACAGAGGAGGATGTTATC 786  
Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
Db 787 ATCATGCTAGTAGGAAATAAACAGATCTTCTGACAGAGCAAGTGTCAATTTGAGGAG 846  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 847 GGAGAGAGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTGCAAAAGCTGA 906  
Qy 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
Db 907 TACAATGTAAAGACGCTCTTTTCGACGTGTAGCAGCAGCTTTGCCGGGAATGGAAGCACA 966  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluThr 200  
Db 967 CAGACAGAACAGAGAAATATGATTGACATATAAACTGGAAGAACCTCAGGAGCAACCA 1026  
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220  
Db 1027 GTCAGTGAAGGAGGCTGTCTCTGCTTAATCTCCCATGTCTTCAACCTTCTTCAGAAAGC 1086  
Qy 221 -----ProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236  
Db 1087 TCACGTCTTTGGCCCCCTTCTCTTTCATTCACGTGACGTGTAATATTTGGCTTGAACCTT 1146  
Qy 237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTyrArg 254  
Db 1147 TTCCCTTCAGTAATAACGATATTGCAATTCATCATTTGCTGCTCTCTCTGAGAGA 1200

RESULT 7  
AAZ93834  
ID AAZ93834 standard; DNA; 719 BP.

XX AAZ93834;  
XX 29-AUG-2000 (first entry)  
XX WTH3 5' partial sequence.  
XX  
KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
KW antibody; immunogen; mutation; detection; therapy; human; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX CDS 49..719  
XX /\*tag= a  
XX /product= "WTH3 C-terminal fragment"

WO200029625-A1.

25-MAY-2000.

18-NOV-1999; 99WO-US027630.

18-NOV-1998; 98US-0108994P.

(NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

Duffy HX, Shan J, Yuan L, Budman D, Calabro A;

WPI; 2000-387828/33.

P-PSDB; AAY83398.

Novel nucleic acids encoding wth3 protein useful for inhibiting multiple  
drug resistance in tumor cells for identifying the modulators of drug  
resistance.

Claim 19; Fig 1a-1b; 82pp; English.

Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)  
Differential Display (MDD) methods and can be used in methods for  
detecting methylation patterns in multiple drug resistance loci. Genes  
are frequently not methylated in cells where they are expressed but are  
methylated in cell types where they are not expressed. Tumour cell DNA is  
often methylated to a different extent and in different regions when  
compared to DNA of normal cells. The methylation pattern in a multiple  
drug resistance locus can be altered and give rise to altered expression  
patterns of that multiple drug resistance locus. Nucleic acids  
corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
can be used as probes for detecting mutations and methylation patterns of  
those loci. The nucleic acids and their homologues are useful for  
inhibition of multiple drug resistance and for treating tumors exhibiting  
multiple drug resistance. They are also useful for detecting and  
measuring the expression of mRNA from identified genes and for  
determining suitable therapeutic treatment. Antibodies directed against  
immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
proteins and polypeptides in tissues or body fluids of patients

SQ Sequence 719 BP; 208 A; 154 C; 193 G; 164 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.2e-121 Length: 719  
Score: 1143.00 Matches: 224  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 87.65% Indels: 0  
DB: 3 Gaps: 0

US-09-441-857-12 (1-254) x AAZ93834 (1-719)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 48 ATGTCGGCGGCGAGACTTCGGGAATTCAGCTGGTGTCTCTGGGG 107

```
QY 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAen 40
Db 108 GAGCAAGCGTGGCAAGACATCTTTGATCACCAGATTGAGTATGACAGATTTCACAAC 167
QY 41 ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 168 ACCTATCAGGCAATTAATGGCAATGACCTTTTATCAAAACATATGTACTTGGAGATGGA 227
QY 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 228 ACAATCGGGCTTCGGCTGGGATACGCGGGTACGAGACGCTCGTAGCCTCATTCCTC 287
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 288 AGGTACATCCGTGATCTGCTGCAGCTGTAGTAGTTTACGATATCACAATGTAACTCA 347
QY 101 PheGlnGlnThrThrTyrTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 348 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGAGAGTGTATGTTATC 407
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 408 ATCAGCTAGTAGGAAATAGAACAGATCTTGCTGACAGAGGCAAGTGTCACTTCAGGAG 467
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 468 GGAGAGAGGAAAGCCAAAGGGCTGAATGTTACGTTTATTGAAACTAGGGCAAAACTGGA 527
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180
Db 528 TACAATGTAAAGCAGCTCTTTCCGCGTGTAGCAGCAGCTTTCCCGGGAATGGAAACACA 587
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200
Db 588 CAGGACGGNAGCAGAGACAGATGAGTGACATATAAACTGGAAAGCCCTCAGGAGCAACA 647
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
Db 648 GTCAGCAAGGGGTGTTCTCTGCTACTCTCCCATGTCTATCTTCAACCCCTCTCAGAAG 707
QY 221 ProProTyrSer 224
Db 708 CCCCTTACTCT 719

RESULT 8
ADL12658
ID ADL12658 standard; cDNA; 3745 BP.
XX
AC ADL12658;
DT
XX
DE 06-MAY-2004 (first entry)
XX
DE Human steroid-induced C3A liver cell cDNA #387.
XX
ss; Gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;
KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
XX
OS Homo sapiens.
XX
PN US6673549-B1.
XX
PD 06-JAN-2004.
XX
PF 12-OCT-2001; 2001US-00976594.
XX
PR 12-OCT-2000; 2000US-0240409P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Furness LM, Buchbinder JL;
XX
DR WPI; 2004-068610/07.
XX
```

```
PT Combination useful for preparing a composition for treating liver
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
PT comprises cDNAs that are differentially expressed in response to steroid
PT treatment.
XX Claim 1; SEQ ID NO 387; 141pp; English.
XX
CC The invention relates to a combination comprising cDNAs that are
CC differentially expressed in response to steroid treatment. Also included
CC are the following: a high throughput method for using a cDNA to detect
CC differential expression of nucleic acids in a sample; and a high
CC throughput method of screening molecules or compounds to identify a
CC ligand that specifically binds a cDNA. The sample is from a subject with
CC Wilson disease and comparison of a standard defines a stage of that
CC disease. The high throughput method of screening molecules or compounds
CC to identify a ligand that specifically binds a cDNA comprises: combining
CC the combination with molecules or compounds under conditions to allow
CC specific binding; and detecting specific binding between each cDNA and at
CC least one molecule or compound. The molecules or compounds are regulatory
CC proteins. The combination is useful for preparing a composition for
CC treating liver disorders associated with steroid therapy, e.g., cirrhosis
CC or hepatitis. The present sequence represents a human cDNA which is
CC differentially expressed in steroid-induced C3A liver cells. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 3745 BP; 1066 A; 755 C; 745 G; 1179 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 2,13e-119 Length: 3745
Score: 1138.00 Matches: 225
Percent Similarity: 89.92% Conservative: 7
Best Local Similarity: 87.21% Mismatches: 22
Query Match: 87.27% Indels: 4
Db: 12 Gaps: 1

US-09-441-857-12 (1-254) x ADL12658 (1-3745)
```

```
QY 1 MetSerAlaGlyLysAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 494 ATGTCACCGGGCGAGACTTCGGGNATCCGTCGAGAAATTCAGCTGGTGTCTCTGGGG 553
QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAen 40
Db 554 GAGCAAGCGTGGAAAGACATCTTTGATCACCAGATTTCATGTATGACAGATTTTCAACAAC 613
QY 41 ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 614 ACCTATCAGGCAAACTTGGCAATTTGACTTTTATCAAAACACTATGTACTTGGAGATCGA 673
QY 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 674 ACAGTACGATTGCAATTAATGGACACAGAGCTCAAGAGCGGTCAGAGAGCTTGATTCCT 733
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 734 AGCTACATTCGTGACTCCACTGTCAGCTGTGTCAGTGTGTTTATGATATATCAAAATGTTAACTCA 793
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 794 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGAGGAGTGTATGTTATC 853
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 854 ATCATGCTAGTAGGAAATATAAACAGATCTTCTGCAAGAGCAAGTGTCAATTCAGGAG 913
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 914 GGAGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTGTCAAAAGCTGGA 973
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180
```

Db	974	TACAAATGTAAGACGAGCTCTTTCCAGCTGTAGCACAGAGCTTTGCCGGGAAATGGAAGACACA	1033
Qy	181	GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr	200
Db	1034	CAGGACAGAGACGAGAGAAGATATGATTGACATAAACTGGAAAAAGCCTTCAGGAGCAACCA	1093
Qy	201	ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys	220
Db	1094	GTCAAGTGAAGAGAGAGCTGTTCCTGCTTAATCTCCCATGTATCTTCAACACTTCTTCAGAAAGC	1153
Qy	221	-----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu	236
Db	1154	TCACCTGCTTTGGCCCCCTTACTCTTTCATTGACTGCAGTGTGAATATTTGCTTGAACTT	1213
Qy	237	PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg	254
Db	1214	TTCCCTTCAGTAATAAAGTATTTGCAATTCATTCATTTGCTGCTGTCTCGTGGAGA	1267

XX	SQ	Sequence 3118 BP; 938 A; 636 C; 638 G; 902 T; 0 U; 4 Other;			
Alignment Scores:					
Pred. No.:	6.12e-119	Length:	3118		
Score:	1133.00	Matches:	224		
Percent Similarity:	89.53%	Conservative:	23		
Best Local Similarity:	86.82%	Mismatches:	7		
Query Match:	86.89%	Indels:	4		
DB:	3	Gaps:	1		
US-09-441-857-12 (1-254) x AAF15989 (1-3118)					
QY	1	MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly	20		
DB	415	ATGTCACGGCGGAGACTTCGGGAATCCGTCAGGAAATCAAGCTGGTGTCTCTGGG	474		
QY	21	GluGlnSerValAlaLysThySerLeuLeuThrArgPheArgTyrAspSerPheAspAsn	40		
DB	475	GAGCAAGCKNTGGAAGACATCTTTGATCACCAGATTTCATGTATGACAGTTTGACAAC	534		
QY	41	ThrTyrGlnAlaIleIleGlyIleAspPheIleSerLysThrMetTyrLeuGluAspGly	60		
DB	535	ACCTATCAGGCACAAATGGCATTTGATCTTTATCAAAAATCTATGTACTTTGGAGGATCGA	594		
QY	61	ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro	80		
DB	595	ACAGTACGATTGCAATTTATGGACACAGCAGGTCAAGAGCGGTTCAGGAGCTTGATTCCT	654		
QY	81	ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer	100		
DB	655	AGCTACATTCGTGACTCCACTGTGGCAGTTGTTGTTTATGATATCAAAATGTTAACTCA	714		
QY	101	PheGlnGlnThrThrLysTyrIleAspAspValArgThrGluArgGlySerAspValIle	120		
DB	715	TTCCAGCAAACTACAAAGTGATGTGATGTGTGAGAACAGAGAGAGAGTGTGTTATC	774		
QY	121	IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu	140		
DB	775	ATCATGTGTAGTGGAAATAAACAGATCTTGTCTGCACAGAGGCAGTGTCAATTGAGGAG	834		
QY	141	GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly	160		
DB	835	GGAGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTATTCAAACTAGTGCAAAAGCTGGA	894		
QY	161	TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr	180		
DB	895	TACAATGTAAGCAGAGCTCTTTCGACGTGTAGCAGCAGCTTTGCGGGGAATGGAAGACA	954		
QY	181	GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr	200		
DB	955	CAGACAGAGACAGAGAAAGATATGATTGAATTAACATGGAAAAGCCTCAGGAGCAACCA	1014		
QY	201	ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys	220		
DB	1015	GTCAAGTGAAGGAGCTGTTCTCTGTAATCTCCCATGTCATCTTCAACCTTCTTCAGAAGC	1074		
QY	221	-----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu	236		
DB	1075	TCACGTCTTTGGGCCCCCTTACTCTTTTCATTTGACTGCAGTGTGAATATGTCCTTGAACCTT	1134		
QY	237	PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTyrArg	254		
DB	1135	TTCCCTTCAGTAATAACGATATGCAATTCATCATTTGCTGCTGCTGCTGCTGCTGCTGCTG	1188		
RESULT 10					
ID	AAZ93836				
XX	AAZ93836 standard; DNA; 739 BP.				
AC	AAZ93836;				
XX					
XX					
DT	29-AUG-2000 (first entry)				
XX					

DE RAB6C coding sequence.  
 XX WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
 KW antibody; immunogen; mutation; detection; therapy; human; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 70..696  
 FT /\*tag= a  
 FT /label= RAB6C polypeptide  
 XX  
 FT  
 XX WO200029625-A1.  
 PN  
 XX  
 PD 25-MAY-2000.  
 XX  
 XX 18-NOV-1999; 99WO-US027630.  
 XX  
 XX 18-NOV-1998; 98US-0108994P.  
 XX  
 XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.  
 PA  
 XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;  
 XX WPI; 2000-387828/33.  
 XX P-PSDB; AAY83400.  
 XX  
 XX Novel nucleic acids encoding wth3 protein useful for inhibiting multiple  
 PT drug resistance in tumor cells for identifying the modulators of drug  
 PT resistance.  
 XX  
 PS Example 1; Fig 3a-3b; 82pp; English.  
 XX  
 CC Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)  
 CC Differential display (MDD) methods and can be used in methods for  
 CC detecting methylation patterns in multiple drug resistance loci. Genes  
 CC are frequently not methylated in cells where they are expressed but are  
 CC methylated in cell types where they are not expressed. Tumour cell DNA is  
 CC often methylated to a different extent and in different regions when  
 CC compared to DNA of normal cells. The methylation pattern in a multiple  
 CC drug resistance locus can be altered and give rise to altered expression  
 CC patterns of that multiple drug resistance locus. Nucleic acids  
 CC corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
 CC can be used as probes for detecting mutations and methylation patterns of  
 CC those loci. The nucleic acids and their homologues are useful for  
 CC inhibition of multiple drug resistance and for treating tumors exhibiting  
 CC multiple drug resistance. They are also useful for detecting and  
 CC measuring the expression of mRNA from identified genes and for  
 CC determining suitable therapeutic treatment. Antibodies directed against  
 CC immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
 CC proteins and polypeptides in tissues or body fluids of patients  
 XX  
 SQ Sequence 739 BP; 219 A; 151 C; 191 G; 178 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: Length: 739  
 Score: 982.00 Matches: 196  
 Percent Similarity: 92.13% Conservatives: 3  
 Best Local Similarity: 90.74% Mismatches: 17  
 Query Match: 75.31% Indels: 0  
 DB: 3 Gaps: 0  
 US-09-441-857-12 (1-254) x AAZ93836 (1-739)  
 QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
 DB 70 ATGTCACGGGGAGACTTCGGGAATCCGTCAGGAAATTCAGCTGGTGTCTCTGGG 129  
 QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgThrAspSerPheAspAsn 40  
 DB 130 GAGCAAGCGTTGGAAAGACATCTTTGATCACCAGATTTCATGACAGTTTTCACCAAC 189  
 QY 41 ThrTyrGlnAlaIleGlyLysAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
 DB 190 ACCTATCAGGCAACAATTGGCAATTGACTTTTATCAAAAACTATGTACTTCGAGGATCGA 249  
 QY 61 ThrIleGlyLeuArgLeuTtpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
 DB 250 ACAATCAGGCTTCAGCTGTGGGATCTCGGGGTGAGAACGTTTCCTAGGCTCATTCCTCC 309  
 QY 81 ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer 100  
 DB 310 AGTTACATCCGTGATTCTGCTGCAGCTGTAGTAGTATACGATATACAAATGTTAACTCA 369  
 QY 101 PheGlnGlnThrThrLysTtpIleAspAspValArgThrGluArgGlySerAspValIle 120  
 DB 370 TTCACGCAAACTACAAAGTGGATTGATGATGTCAGAACAGAAAGGAAGTGTATTATC 429  
 QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
 DB 430 ATCATGCTAGTAGGAATATAACAGATCTTGCTCACAAGAGGCAAGTGTCAATTGAGGAG 489  
 QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
 DB 490 GGAGAGAGGAAAGCAAGAGCTGAATGTTATGTTTATTGAAACTAGTGCAGAAAGCTGGA 549  
 QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLalaLalaLeuProGlyMetGluSerThr 180  
 DB 550 TACAATGTAAAGCAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCCGCGGAATGGAAGCACA 609  
 QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
 DB 610 CAGACAGAGACAGAGAGATATGATTGACATAAACTGGANAAGCCCTCAGAGCAACCA 669  
 QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThr 216  
 DB 670 GTCAGTGAAGGAGGCTGTCTCTGCTAATCTCCCATGTCATCTTCAACC 717  
 RESULT 11  
 AAZ93835  
 ID AAZ93835 standard; DNA; 718 BP.  
 AC AAZ93835;  
 XX 29-AUG-2000 (first entry)  
 DE RAB6 coding sequence.  
 KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;  
 KW antibody; immunogen; mutation; detection; therapy; human; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 49..675  
 FT /\*tag= a  
 FT /product= "RAB6"  
 XX  
 PN WO200029625-A1.  
 XX  
 PD 25-MAY-2000.  
 XX  
 XX 18-NOV-1999; 99WO-US027630.  
 XX  
 XX 18-NOV-1998; 98US-0108994P.  
 XX  
 XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.  
 PA  
 XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;  
 XX WPI; 2000-387828/33.  
 XX P-PSDB; AAY83399.  
 XX  
 XX Novel nucleic acids encoding wth3 protein useful for inhibiting multiple  
 PT drug resistance in tumor cells for identifying the modulators of drug  
 PT resistance.

XX

PS Example 1; Fig 1a-1b; 82pp; English.

CC Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)  
CC Differential Display (MDD) methods and can be used in methods for  
CC detecting methylation patterns in multiple drug resistance loci. Genes  
CC are frequently not methylated in cells where they are expressed but are  
CC methylated in cell types where they are not expressed. Tumour cell DNA is  
CC often methylated to a different extent and in different regions when  
CC compared to DNA of normal cells. The methylation pattern in a multiple  
CC drug resistance locus can be altered and give rise to altered expression  
CC patterns of that multiple drug resistance locus. Nucleic acids  
CC corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
CC can be used as probes for detecting mutations and methylation patterns of  
CC those loci. The nucleic acids and their homologues are useful for  
CC inhibition of multiple drug resistance and for treating tumors exhibiting  
CC multiple drug resistance. They are also useful for detecting and  
CC measuring the expression of mRNA from identified genes and for  
CC determining suitable therapeutic treatment. Antibodies directed against  
CC immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
CC proteins and polypeptides in tissues or body fluids of patients  
XX

SQ Sequence 718 BP; 219 A; 141 C; 184 G; 174 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6.31e-100 Length: 718  
Score: 959.00 Matches: 191  
Percent Similarity: 90.28% Conservative: 4  
Best Local Similarity: 88.43% Mismatches: 21  
Query Match: 73.54% Indels: 0  
DB: 3 Gaps: 0

US-09-441-857-12 (1-254) x AAZ93835 (1-718)

QY 1 MetSerAlaGlyArgPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
DB 49 ATGTCACGGCGGAGACTTCGGGAATCCGCTGAGGAATTCAGCTGGTCTCTCGGG 108  
QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgPheArgPheAspAsn 40  
DB 109 GAGCAAAACGTTGGAAACACATCTTTGATCACCAGATTTCATGATGACAGTTTTCACAA 168  
QY 41 ThrTyrglnAlaIleGlyLeuAspPheLeuSerLysThrMetTyrglnLeuGluAspGly 60  
DB 169 ACCTATCAGGACAAATTCGATGATCTTTTATCAAAAATATGCTACTCGAGATCGA 228  
QY 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
DB 229 ACAGTACGATTGCAATTATGGGACACACAGCAGTCAAGAGCGGTCAGGAGCTTGATTCCT 288  
QY 81 ArgTyrlleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
DB 289 AGCTACATTCGTGACTCCACTGCGCAGTTCTTGTATGATATACAAAATGTTAACTCA 348  
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120  
DB 349 TTCAGAGAACTACAAAGTGGATGATGATGTCAGAACAGAAAGAGGAAGTGTATC 408  
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
DB 409 ATCATGCTAGTAGGAATAAACACAGATCTTCTGACAAAGAGCGAAGTCAATTGAGGAG 468  
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
DB 469 GGAGAGAGAAAGCAAGAGCTGAATGTTATGTTTATTGAAACTAGTGTCAAAAAGCTGGA 528  
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180  
DB 529 TACAAATGTAAAGCAGCTCTTTTCGACGTGTACAGCAGCTTTTCCCGGAATGGAAGACCA 588  
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluThr 200  
DB 589 CAGCACAGAGCAGAGAAGATATGATTGACATATAAACTGGAAAGCCTCAGGACCAACCA 648

QY 201 ValSerGluGlyCysSerCysTyrsSerProMetSerSerSerThr 216  
DB 649 GTCAGTGAAGAGGCTGTTCTCTGTAATGTCCTAGTCACTTCAACC 696

## RESULT 12

AAZ93837  
ID AAZ93837 standard; DNA; 740 BP.

XX AAZ93837;

AC AAZ93837;

DT 29-AUG-2000 (first entry)

XX RAB6 coding sequence.

DE WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;

XX antibody; immunogen; mutation; detection; therapy; human; ss.

KW Homo sapiens.

XX Key Location/Qualifiers

FT CDS 71..697

FT /\*tag= a

FT /product= "RAB6"

XX WO200029625-A1.

XX 25-MAY-2000.

XX 18-NOV-1999; 99WO-US027630.

XX 18-NOV-1998; 98US-0108994P.

XX (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.

XX Duffy HX, Shan J, Yuan L, Budman D, Calabro A;

XX WPI; 2000-387828/33.

XX P-PSDB; AAY83399.

XX Novel nucleic acids encoding wth3 protein useful for inhibiting multiple

XX drug resistance in tumor cells for identifying the modulators of drug

XX resistance.

XX Example 1; Fig 3a-3b; 82pp; English.

XX Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)  
XX Differential Display (MDD) methods and can be used in methods for  
XX detecting methylation patterns in multiple drug resistance loci. Genes  
XX are frequently not methylated in cells where they are expressed but are  
XX methylated in cell types where they are not expressed. Tumour cell DNA is  
XX compared to DNA of normal cells. The methylation pattern in a multiple  
XX drug resistance locus can be altered and give rise to altered expression  
XX patterns of that multiple drug resistance locus. Nucleic acids  
XX corresponding to the identified loci such as WTH3 and RAB6 nucleic acids  
XX can be used as probes for detecting mutations and methylation patterns of  
XX those loci. The nucleic acids and their homologues are useful for  
XX inhibition of multiple drug resistance and for treating tumors exhibiting  
XX multiple drug resistance. They are also useful for detecting and  
XX measuring the expression of mRNA from identified genes and for  
XX determining suitable therapeutic treatment. Antibodies directed against  
XX immunogenic fragments of WTH3 and RAB6 are useful for detecting specific  
XX proteins and polypeptides in tissues or body fluids of patients  
XX

SQ Sequence 740 BP; 223 A; 146 C; 193 G; 178 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6.59e-100 Length: 740  
Score: 959.00 Matches: 191  
Percent Similarity: 90.28% Conservative: 4  
Best Local Similarity: 88.43% Mismatches: 21  
Query Match: 73.54% Indels: 0

```
DB: 3 Gaps: 0
US-09-441-857-12 (1-254) x AAZ93837 (1-740)
QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 71 ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 130
QY 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40
Db 131 GAGCAAGCGTGGGAAGACATCTTTGATCACACAGATTATGTATGACAGATTGTGACAC 190
QY 41 ThrTyrGlnAlaLeuLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 191 ACCTATCAGCAACAATTCGCTTGTCTTTATCAAAAATATGTACTTGGAGGATCGA 250
QY 61 ThrLeuGlyLeuArgLeuThrAspThrAlaGlyGlnGluArgLeuSerLeuLeuPro 80
Db 251 ACAGTACGATTGCAATATATGGGACACAGAGGTTCAGAGCGGTTCAGGAGCTTGATTCCT 310
QY 81 ArgTyrIleArgAspSerAlaAlaValValValValTyrAspIleThrAsnValAsnSer 100
Db 311 AGCTACATTCGTGACTCCACTGTGGCAGTTGTTTATGATATCAAAATGTAACTCA 370
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 371 TTCCAGCAAACTACAAAGTGGATTGATGTGATGTCAGAACAGAAAGAGGAAGTGTATC 430
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 431 ATCATGCTAGTAGGAATAAACAAGATCTTGTCTGACAGAGCAAGTGTCAATTTGAGGAG 490
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 491 GGAGAGAGGAAGCCAAAGAGCTGAATGTATGTTTATTTGAAACTAGTCAAAAGCTGA 550
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180
Db 551 TACAATGTAAAGCAGCTCTTCGACGTGTAGCAGCAGCTTTCGCGGAATGGAAGCAC 610
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200
Db 611 CAGGACAGAAAGCAGAGAGATATGATTGACATAAAACTCGAAAGCCCTCAGGAGCAACCA 670
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThr 216
Db 671 CTCAGTAGAGGAGGCTGTTCTCTGCTAATGTCCTCTAGTCACTCTCAACC 718
RESULT 13
ID AAZ93841
XX AAZ93841 standard; DNA; 624 BP.
AC AAZ93841;
XX 29-AUG-2000 (first entry)
DE RAB6C coding sequence.
KW WTH3; RAB6; drug resistance; locus; loci; methylation; tumour; probe;
KW antibody; immunogen; mutation; detection; therapy; human; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..624
FT FT /*tag= a
FT FT /product= "RAB6C"
XX WO200029625-A1.
PN 25-MAY-2000.
PD 18-NOV-1999; 99WO-US027630.
PF
```

```
XX 18-NOV-1998; 98US-0108994P.
PR (NSHO-) NORTH SHORE-LONG ISLAND JEWISH RES.
PA Duffy HX, Shan J, Yuan L, Budman D, Calabro A;
PI WPI; 2000-387828/33.
XX P-PSDB; AAY83403.
PT Novel nucleic acids encoding wth3 protein useful for inhibiting multiple
PT drug resistance in tumor cells for identifying the modulators of drug
XX resistance.
PS Claim 6; Page 71-72; 82pp; English.
XX Nucleic acids encoding WTH3 and RAB6 were isolated by Methyl- (or mutant)
CC Differential Display (MDD) methods and can be used in methods for
CC detecting methylation patterns in multiple drug resistance loci. Genes
CC are frequently not methylated in cells where they are expressed but are
CC methylated in cell types where they are not expressed. Tumour cell DNA is
CC often methylated to a different extent and in different regions when
CC compared to DNA of normal cells. The methylation pattern in a multiple
CC drug resistance locus can be altered and give rise to altered expression
CC patterns of that multiple drug resistance locus. Nucleic acids
CC corresponding to the identified loci such as WTH3 and RAB6 nucleic acids
CC can be used as probes for detecting mutations and methylation patterns of
CC those loci. The nucleic acids and their homologues are useful for
CC inhibition of multiple drug resistance and for treating tumors exhibiting
CC multiple drug resistance. They are also useful for detecting and
CC measuring the expression of mRNA from identified genes and for
CC determining suitable therapeutic treatment. Antibodies directed against
CC immunogenic fragments of WTH3 and RAB6 are useful for detecting specific
CC proteins and polypeptides in tissues or body fluids of patients
XX SQ Sequence 624 BP; 199 A; 115 C; 162 G; 148 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,51e-99 Length: 624
Score: 953.00 Matches: 189
Percent Similarity: 92.31% Conservative: 3
Best Local Similarity: 90.87% Mismatches: 16
Query Match: 73.08% Indels: 0
DB: 3 Gaps: 0
US-09-441-857-12 (1-254) x AAZ93841 (1-624)
QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 1 ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60
QY 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40
Db 61 GAGCAAGCGTGGGAAGACATCTTTGATCACACAGATTATGTATGACAGATTGTGACAC 120
QY 41 ThrTyrGlnAlaLeuLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 121 ACCTATCAGGCAACAATTTGGCATTGACTTTTATCAAAAATATGTACTTGGAGGATCGA 180
QY 61 ThrLeuGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuLeuPro 80
Db 181 ACAATCAGGCTTCAGCTGTGGGATCTCGCGGTGAGGAACGTTTCCGTAGCTCATTCCTCC 240
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 241 AGTTACATCCCGTGATTTCTGCTGACGCTGTAGTAGTTACGATATACAAAATGTTAACTCA 300
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 301 TTCAGCAAACTACAAAGTGGATTGATGTGATGTCAGAACAGAAAGAGGAGGTGATGTATC 360
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
PF
```



Db 361 ATCATGCTAGTAGGAAATAAACAGATCTTGCTGACAGAGCAAGTGTCAATTGAGGAG 420  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 421 GGAGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAACTAGTGTCAAAAGCTGGA 480  
Qy 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180  
Db 481 TACAATGTAAAGCAGCTCTTCGAGCTGTAGCAGCAGCTTTGCCGGGAATGGAAGCACAA 540  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
Db 541 CAGGACAGAAGCAGAGAAGATGATTGACATAAACTGGAAGAAAGCCTCAGGAGCAACA 600  
Qy 201 ValSerGluGlyGlyCysSerCys 208  
Db 601 GTCAAGTGAAGGAGGCTGTTCTCTGC 624

## RESULT 14

AAK52711

ID AAK52711 standard; cDNA; 1219 BP.

XX AC

XX AAK52711;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 2240.

XX DE

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.

XX OS

XX Homo sapiens.

XX PN

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX PA

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang Z;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX P-PSDB; AAM79578.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

XX in diagnosis and gene therapy.

XX Claim 1; Page 4593; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to

XX cytokine, cell proliferation or cell differentiation or which may induce

XX production of other cytokines in other cell populations. The

XX polynucleotides and polypeptides are useful in gene therapy, vaccines or

XX peptide therapy. The polypeptides have various cytokine-like activities,

XX e.g. stem cell growth factor activity, haematopoiesis regulating

XX activity, tissue growth factor activity, immunomodulatory activity and

XX activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication

XX SQ Sequence 1219 BP; 287 A; 314 C; 337 G; 281 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6,48e-92 Length: 1219  
Score: 892.00 Matches: 172  
Percent Similarity: 90.38% Conservative: 16  
Best Local Similarity: 82.69% Mismatches: 20  
Query Match: 68.40% Indels: 0  
DB: 4 Gaps: 0

US-09-441-857-12 (1-254) x AAK52711 (1-1219)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 179 ATGTCGCGAGGGGAGATTTTGGGAATCCACTGAGAAAATTCAGTTGGTGTCTTGGGG 238  
Qy 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40  
Db 239 GAGCAGAGCGTCGGGAAGACGCTCTCTGATTACGAGGTTTCATGTACGACAGCTTCGACAAC 298  
Qy 41 ThrTyrGlnAlaIleIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 299 ACATACCAGGCAACCATTTGGGATTGACTTCTTGTCAAAAACCATGTACTTGGAGGCGC 358  
Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
Db 359 ACGTGGCAGTGCAGCTCTGGGACACAGCTGTCAGGAGAGTTCCGAGAGCTGATCCCC 418  
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
Db 419 AGCTACATCCGGGACTCCACGGTGGCTGGTGGTGTACGACATCACAAATCTCAACTCC 478  
Qy 101 PheGlnGlnThrThrLysTyrIleAspAspValArgThrGluArgGlySerAspValIle 120  
Db 479 TTCCAACAGACCTCTAAGTGGATCGACGAGCTCAGAGAGGGGGCAGTGTATTATC 538  
Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
Db 539 ATCATGCTGGTGGCAACAGACGGACCTGCTGATAGAGGCGAGATAACCATCGAGAG 598  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 599 GGGGAGCAGCGCGCCAAAGAACTGAGCGTCTATGTTTCATTGAGACCAGTGCAGAGACTGGC 658  
Qy 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
Db 659 TACAACGTGAAGCAGCTTTTTCGACGTGGCGCTCTACCCGGAATGGGAATGTC 718  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
Db 719 CAGGAGAAAGCAAGAGGAGATGATTGACATCAAGCTGGACAAACCCAGGAGCCCCCG 778  
Qy 201 ValSerGluGlyGlyCysSerCys 208  
Db 779 GCCAGCGAGGGCGCTGCTCTCTGC 802

## RESULT 15

ADP07320

ID ADP07320 standard; DNA; 1266 BP.

XX AC

XX ADP07320;

XX 29-JUL-2004 (first entry)

XX Human RAB6B DNA.

XX ds; proliferative disease; breast; methylation; CpG; bisulfite; human.



OS Homo sapiens.  
 XX DE10255104-A1.  
 XX 11-MAR-2004.  
 XX  
 XX 26-NOV-2002; 2002DE-01055104.  
 XX 27-AUG-2002; 2002DE-01039313.  
 XX (EPITG-) EPIGENOMICS AG.  
 XX Maier S;  
 XX WPI: 2004-284340/27.  
 XX GENBANK; NM\_016577.  
 XX  
 XX Analyzing proliferative diseases of breast cells, useful e.g. for  
 PT diagnosis, prognosis and treatment of breast cancer, by determining  
 PT methylation status of specific genes.  
 XX  
 XX Claim 1; Page; 22pp; German.  
 XX  
 CC This invention describes a novel method of analysing proliferative  
 CC diseases of breast cells by determining the methylation status of certain  
 CC genes. The invention also describes nucleic acids or their complements,  
 CC oligomers, especially oligonucleotides or peptide nucleic acid oligomers,  
 CC that hybridise to, or are identical with, any of the nucleic acids, the  
 CC preparation of an oligomer array for analysing proliferative diseases of  
 CC breast cells that are associated with the methylation status of CpG  
 CC dinucleotides of the genes by bonding at least one oligomer to a solid  
 CC phase and a kit comprising a bisulfite reagent and the oligomer. The  
 CC nucleic acids are genomic sequences (5'- and/or regulatory and/or CpG-  
 CC rich regions). The base sequence of the oligomer includes at least one  
 CC CpG island, especially with C in the middle third of the sequence. The  
 CC process involves a genomic DNA sample treated chemically, specifically  
 CC with a bisulfite reagent, to convert non-methylated C to uracil or some  
 CC other base having base-pairing properties different from those of C.  
 CC Fragments of the treated DNA are then amplified, using primers and a  
 CC polymerase and the methylation status of the genomic CpG dinucleotides is  
 CC determined by analysis of the amplicons, particularly by hybridisation to  
 CC the oligomer, optionally with extension of the hybridised oligomer by at  
 CC least one base, or detection is by sequencing. The amplification may use  
 CC methylation-specific primers. Alternatively, genomic DNA is extracted  
 CC from a sample and digested with methylation-specific restriction enzymes,  
 CC then the digestion fragments detected, optionally after amplification. In  
 CC either method, more than 10 fragments of 100-200 bp are amplified in a  
 CC single reaction vessel, using a heat-resistant DNA polymerase in PCR. The  
 CC amplicons carry detectable markers, e.g. fluorophores, radioisotopes  
 CC and/or releasable fragments of known mass that can be detected by mass  
 CC spectrometry. The method is used for characterisation, classification,  
 CC differentiation, staging, phase-estimation, diagnosis and/or therapy of  
 CC proliferative diseases of breast cells. The method provides very specific  
 CC classification of proliferative diseases, allowing better treatment. It  
 CC can both characterise methylation status and detect single-nucleotide  
 CC polymorphisms. This sequence represents human gene used to illustrate the  
 CC method of the invention. NOTE: This sequence does not appear in the  
 CC printed specification but has been retrieved from Genbank.  
 XX  
 SQ Sequence 1266 BP; 269 A; 379 C; 344 G; 274 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6,84e-92 Length: 1266  
 Score: 892.00 Matches: 172  
 Percent Similarity: 90.38% Conservatives: 16  
 Best Local Similarity: 82.69% Mismatches: 20  
 Query Match: 68.40% Indels: 0  
 DB: 12 Gaps: 0

US-09-441-857-12 (1-254) x ADP07320 (1-1266)

QY 1 MetSerAlaGlyCysPheGlyAspPheGlyAsnProLeuArgLysPheLeuValPheLeuGly 20  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 338 ATGTCGCGAGGGGAGATTTTGGGAATCCACTGAGAAAATTCAGTTGGTGTCTTGGGG 397  
 QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAen 40  
 Db 398 GAGCAGACGCGTGGGAGACGCTCTTGATTAGAGGTTTCATGTACGACAGCTTCGACAAAC 457  
 QY 41 ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
 Db 458 ACATACCAGGCAACCATTGGGATTGACTTCTTGTCAAAAACCATGTACTTGGAGGACCGC 517  
 QY 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
 Db 518 ACGGTGCGACTGCGACACTCTGGGACACAGCTGGTCAGGAGAGGTTCGCGAGCTGTATCCCC 577  
 QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
 Db 578 AGCTACATCCGGGACTCCACGGT 637  
 QY 101 PheGlnGlnThrThrLysThrLysThrLysThrLysThrLysThrLysThrLysThrLys 120  
 Db 638 TTCCAACAGACCTCTTAAGTGGATCGACGACGTTCAGGACAGAGAGGGGCGAGTGTATC 697  
 QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
 Db 698 ATCATGTGTGTGGGCAACAAGACGACCTGGCTGATAAGAGGCGACATAACCATCGAGGAG 757  
 QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
 Db 758 GGGGAGCAGCGCGCCCAAGAACTGAGCGTCTATGTTGAGACGACGATGCGAAGACTGGC 817  
 QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
 Db 818 TACAACGTGAAGCAGCTTTTTCGACGTGTGGCGTCTTACCCGGAATGGAGAATGTC 877  
 QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
 Db 878 CAGGAGAAAAGCAAGAAGGGATGATCGACATCAAGCTGGACAAACCCAGGAGCCCCCG 937  
 QY 201 ValSerGluGlyGlyCysSerCys 208  
 Db 938 GCCAGCAGGGCGGCTGTCTCTGTC 961

Search completed: April 25, 2005, 03:56:04

Job time : 609 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 25, 2005, 03:04:50 ; Search time 170 Seconds  
(without alignments)  
2444.790 Million cell updates/sec

Title: US-09-441-857-12  
Perfect score: 1304  
Sequence: 1 MSAGDGFQNLKFKLVLG.....NLFPSLITFCNSLLPLVSWR 254

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-MODEL=frame+ p2n.model -DEV-xlp  
-Q/cgn2\_1/USFTO\_spool\_p/US09441857/runat\_22042005\_155509\_24777/app.query.fasta\_1.391  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09441857 @CGN 1 1.105 @runat\_22042005\_155509\_24777 -NCFU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

- 1: /cgn2\_6/prodata/1/ina/5A COMB.seq:
- 2: /cgn2\_6/prodata/1/ina/5B COMB.seq:
- 3: /cgn2\_6/prodata/1/ina/6A COMB.seq:
- 4: /cgn2\_6/prodata/1/ina/6B COMB.seq:
- 5: /cgn2\_6/prodata/1/ina/PCITUS COMB.seq:
- 6: /cgn2\_6/prodata/1/ina/backfilee1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1138	87.3	3745	4	US-09-976-594-387
2	973	74.6	739	4	US-09-949-016-2946
3	479	36.7	329	4	US-09-513-999C-2679
4	401.5	30.8	414	4	US-09-248-796A-6213
5	396	30.4	1546	4	US-09-949-016-5540
6	391.5	30.0	719	4	US-09-949-016-3023
7	388.5	29.8	1984	4	US-09-023-655-7
8	377	28.9	2874	4	US-09-949-016-2160
9	371.5	28.5	1630	4	US-09-949-016-409
10	354.5	27.2	921	4	US-09-016-434-1124
11	347.5	26.6	773	4	US-09-949-016-4216
12	346.5	26.6	717	4	US-09-248-796A-5859

13	344	26.4	834	4	US-09-949-016-3883	Sequence 3883, Ap
14	342.5	26.3	1631	4	US-09-620-312D-587	Sequence 587, Ap
c 15	338	25.9	998	4	US-09-270-767-1882	Sequence 1882, Ap
c 16	338	25.9	998	4	US-09-270-767-17164	Sequence 17164, A
17	337.5	25.9	714	4	US-09-248-796A-5886	Sequence 5886, Ap
18	337	25.8	723	4	US-09-016-434-1422	Sequence 1422, Ap
19	335.5	25.7	1775	4	US-09-949-016-4926	Sequence 4926, Ap
20	335	25.7	1069	4	US-09-620-312D-646	Sequence 646, Ap
21	333	25.5	1255	4	US-09-949-016-1772	Sequence 1772, Ap
22	333	25.5	8137	4	US-09-566-921-7	Sequence 7, Appli
23	331.5	25.4	848	3	US-08-741-411-2	Sequence 2, Appli
24	331.5	25.4	920	4	US-09-949-016-4287	Sequence 4287, Ap
25	328.5	25.2	833	4	US-09-620-312D-426	Sequence 426, Ap
26	325.5	25.0	1053	4	US-09-270-767-12052	Sequence 12052, A
27	322.5	24.7	875	3	US-09-075-454-10	Sequence 10, Appl
28	322.5	24.7	1106	4	US-09-620-312D-959	Sequence 959, App
29	322.5	24.7	2612	3	US-09-484-970B-142	Sequence 142, App
30	322	24.7	639	3	US-09-399-913-66	Sequence 66, Appl
31	322	24.7	639	4	US-09-350-614-66	Sequence 66, Appl
32	321.5	24.7	809	4	US-09-949-016-3376	Sequence 3376, Ap
33	321.5	24.7	1148	4	US-09-949-016-4879	Sequence 4879, Ap
c 34	321.5	24.7	24257	4	US-09-949-016-13902	Sequence 13902, A
35	318.5	24.4	925	2	US-08-916-901-4	Sequence 4, Appli
36	318.5	24.4	925	3	US-09-154-602-4	Sequence 4, Appli
37	316.5	24.3	2552	4	US-09-270-767-15247	Sequence 15247, A
38	314.5	24.1	842	4	US-09-255-920A-6	Sequence 6, Appli
39	309.5	23.7	3936	4	US-09-919-172-49	Sequence 49, Appl
40	306.5	23.5	970	3	US-08-888-077A-28	Sequence 28, Appl
41	306.5	23.5	2334	4	US-09-949-016-1758	Sequence 1758, Ap
42	306	23.5	1275	4	US-09-949-016-3508	Sequence 3508, Ap
43	304.5	23.5	1799	4	US-09-774-528-36	Sequence 36, Appl
44	304.5	23.4	684	4	US-09-248-796A-6188	Sequence 6188, Ap
45	303	23.2	1511	4	US-09-949-016-1134	Sequence 1134, Ap

ALIGNMENTS

RESULT 1  
US-09-976-594-387  
; Sequence 387, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 387  
; LENGTH: 3745  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; OTHER INFORMATION: Incyte ID No. 6673549 107569.15  
US-09-976-594-387

Alignment Scores:	1.9e-134	Length:	3745
Pred. No.:	1138.00	Matches:	225
Score:	89.21%	Conservative:	7
Best Local Similarity:	87.32%	Mismatches:	22
Query Match:	87.27%	Indels:	4
DB:	4	Gaps:	1

US-09-441-857-12 (1-254) x US-09-976-594-387 (1-3745)

Oy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
|||||



```
; SEQ ID NO 2679
; LENGTH: 329
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..329
US-09-513-999C-2679

Alignment Scores:
Pred. No.: 6,47e-52 Length: 329
Score: 479.00 Matches: 92
Percent Similarity: 91.43% Conservative: 4
Best Local Similarity: 87.62% Mismatches: 9
Query Match: 36.73% Indels: 0
DB: 4 Gaps: 0

US-09-441-857-12 (1-254) x US-09-513-999C-2679 (1-329)
Qy 5 GlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGlyGlnSerVal 24
Db 12 GGAGATTGGAATCCACTGAGAAAATTCAAGTTGGTCTTGGGGGAGCAGCGTC 71
Qy 25 AlalysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAla 44
Db 72 GGGAGACGCTCTCGATTACGAGGTTTCATGTACGACAGCTTCGACACACATACCAGCA 131
Qy 45 IleIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeu 64
Db 132 ACCATTGGGATTGACTCTTGTCAAAAACCATGACTTGGAGGACCGCAGCGTGACTG 191
Qy 65 ArgLeuTpAspThrAlaGlyGlnGluArgSerLeuIleProArgTyrIleArg 84
Db 192 CAGCTCTGGGACACAGCTGTCAGAGAGGTTCCGACGCTGATCCCGACTACATCCG 251
Qy 85 AspSerAlaAlaValValValTyrAspIleThrAsnValAsnSerPheGlnGlnThr 104
Db 252 GACTCCACGGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 311
Qy 105 ThrLysTpIleAsp 109
Db 312 TCTAAGTGGATCGAC 326

RESULT 4
US-09-248-796A-6213
; Sequence 6213, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6213
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-6213

Alignment Scores:
Pred. No.: 7.57e-42 Length: 414
Score: 401.50 Matches: 80
Percent Similarity: 79.46% Conservative: 9
Best Local Similarity: 71.43% Mismatches: 22
Query Match: 30.79% Indels: 1
DB: 4 Gaps: 1

US-09-441-857-12 (1-254) x US-09-248-796A-6213 (1-414)
Qy 6 AspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGlyGlnSerValAla 25
Db 79 GATAAATCAATTTTACTTAAATAATACAGATTGCTTTCTTGGTGATCAAGTGTGGT 138
Qy 26 LysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIle 45
Db 139 AAAACATCAATTAATCACCAGATTATGTATGATGATCAATTTGATGAACATTTATGCTGCCACG 198
Qy 46 IleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeu 64
Db 199 ATTGGAATTGATTTTTTATCGAAACCAATGATTATTAGAAAGAGGTAAACCATTAGATTA 258
Qy 65 ArgLeuTpAspThrAlaGlyGlnGluArgSerLeuIleProArgTyrIleArg 84
Db 259 CAATTATGGGATCTGCGGACAGAAAGATTTTCGATCAATTAATACCTTCATATATTAGA 318
Qy 85 AspSerAlaAlaValValValTyrAspIleThrAsnValAsnSerPheGlnGlnThr 104
Db 319 GATTCTCATGTTCAGTAATATGTTATGATATAACCAATAAAAAATCATTGATAATCTT 378
Qy 105 ThrLysTpIleAspAspValArgThrGluArgGly 116
Db 379 GATAAATGGATTAAAGATGTTTAAATTAGAACGAGGT 414

RESULT 5
US-09-949-016-5540
; Sequence 5540, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5540
; LENGTH: 1546
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5540

Alignment Scores:
Pred. No.: 3.22e-40 Length: 1546
Score: 396.00 Matches: 83
Percent Similarity: 57.80% Conservative: 43
Best Local Similarity: 38.07% Mismatches: 86
Query Match: 30.37% Indels: 6
DB: 4 Gaps: 2

US-09-441-857-12 (1-254) x US-09-949-016-5540 (1-1546)
Qy 8 GlyAsnProLeuArgLysPheLysLeuValPheLeuGlyGlnSerValAlaLysThr 27
Db 134 GGGAAACAAGATCTGTCATTTAAGCTGTTCTGCTGGGGAGTCTCGGTAGGCAATCC 193
Qy 28 SerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleGly 47
Db 194 AGCCTCGCTCCGCTTTGTCAAGGACAGTTTCCAGTACACGAGAGACCAATTCGA 253
Qy 48 IleAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTrp 67
Db 254 GCGGCTTCTCTCACAGACTGTCTGCTGTGATGACACACAGTCAAGTCTTGAGATCTGG 313
```



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1984 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: U937NOT01  
CLONE: 000513  
US-09-023-655-7

## Alignment Scores:

Pred. No.: 4,38e-39 Length: 1984  
Score: 388.50 Matches: 82  
Percent Similarity: 58.8% Conservatives: 41  
Best Local Similarity: 39.23% Mismatches: 81  
Query Match: 29.79% Indels: 5  
DB: 4 Gaps: 2

US-09-441-857-12 (1-254) x US-09-023-655-7 (1-1984)

Qy 8 GlyAsnProLeuArgGlyPheLysLeuValPheLeuGlyGluGlnSerValAlaLysThr 27  
Db 11 GGAAATAAAATATGCGAGTTCAAACTAGTACTTCTGGGAGAGTCCGCTGTTGGCAATCA 70  
Qy 28 SerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGly 47  
Db 71 AGCCTAGTGCCTTCGTTTGTGAAGGCCCAATTCATGAATTTCAAGAGAGTACCATTCGGG 130  
Qy 48 IleAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrp 67  
Db 131 GCTGCTTTTCAACCAACTGTATGCTCTGTATGACACTACAGTAAAGTTTCAATATGG 190  
Qy 68 AspThrAlaGlyGlnGluArgSerLeuIleProArgTyrIleArgAspSerAla 87  
Db 191 GATACAGCTGGTCAAGACGATACCATAGCTAGCACCAATGTACTACAGAGGACGACAA 250  
Qy 88 AlaAlaValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrThrLysTyrp 107  
Db 251 GCAGCCATAGTTGTATATGATATCAAAATGAGGAGTCTCTTTCAGAGGACCAAAAAATTGG 310  
Qy 108 IleAspAspValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArg 127  
Db 311 GTTAAAGAACTTCAGGCAACCAAGTCTCTACATTTGTAATAGCTTTATCGGGAACAAAG 370  
Qy 128 ThrAspLeuAlaAspLysArgGlnValSerValGluGlyGluArgLysAlaLysGly 147  
Db 371 GCCGACCTAGCAAAATAAAGACGAGTATGATTTCCAGGAAGCACAGTCTCTATCGAGATGAC 430  
Qy 148 LeuAsnValThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuphe 167  
Db 431 AATAGTTTATTATTTCATGGAGACATCCGCTAAACATCAATGAATGTAATGAATATATTC 490  
Qy 168 ArgArgValAlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAsp 187  
Db 491 ATGGCAATAGCTAAAAAATTGCCAAGAAATGAACCAAAATCCAGGAGCAAAATTCCTGCC 550  
Qy 188 MetSer---AspIleLysLeuGlyProGlnGlnThrValSerGluGlyCys 206  
Db 551 AGAGGAAGGAGGTAGACCTTACCGAACCACACAAACCAACGAGGAATCAG----- 601  
Qy 207 SerCysTyrSerProMetSerSerSer 215  
Db 602 ---TGTGTGTAGTAACCTTCTAGT 625

## RESULT 8

US-09-949-016-2160  
Sequence 2160, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIORITY FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIORITY FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIORITY FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIORITY FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2160  
LENGTH: 2874  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-2160

## Alignment Scores:

Pred. No.: 2,35e-37 Length: 2874  
Score: 377.00 Matches: 91  
Percent Similarity: 49.44% Conservatives: 41  
Best Local Similarity: 34.08% Mismatches: 105  
Query Match: 28.91% Indels: 30  
DB: 4 Gaps: 4

US-09-441-857-12 (1-254) x US-09-949-016-2160 (1-2874)

Qy 13 LysPheLysLeuValPheLeuGlyGluGlnSerValAlaLysThrSerLeuIleThrArg 32  
Db 141 CAGTTCAAATTTGGTCTCTGCGAGAAATCTCGAGTGGGAAAGTCAAGCTCGTATTACGT 200  
Qy 33 PheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLysPheLeuSer 52  
Db 201 TTTGTCAAGGGCGAGTCCATGAGTACAGAGAGACACCATTTGGAGCGGCTTCTCACC 260  
Qy 53 LysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrAspThrAlaGlyGln 72  
Db 261 CAGTCCGTTTCTAGATGACACACAGTGAAGTTTGAGATCTGGGACACAGCTGGCAG 320  
Qy 73 GluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaValVal 92  
Db 321 GAGCGATATCACAGCTTAGCCCTTACATGATCTACAGGGGTGCCCAAGCTGCAATCGTGT 380  
Qy 93 TyrAspIleThrAsnValAsnSerPheGlnGlnThrThrLysTyrIleAspValArg 112  
Db 381 TACGACATTACTAATCAGNAACCTTTGCCGAGCAAGACATGGGTGAAGAACTACAG 440  
Qy 113 ThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeuAlaAsp 132  
Db 441 CGACAGGCGAGTCTAGCATCGTTATTCCTCTGGCAGGGAACAAAGCTGACCTGGCCAC 500  
Qy 133 LysArgGlnValSerValGluGlyGluArgLysAlaLysGlyLeuAsnValThrPhe 152  
Db 501 AAACGTATGTGGAGTATGAAGAGCCCGAGCATATGCAGATGCACACAGCTTATTGTTTC 560  
Qy 153 IleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgValAlaAla 172  
Db 561 ATGGAGACTTCAGCCCAAGACAGCTATGAACGTGAATGATCTCTCTCGCAATAGCTAAG 620  
Qy 173 AlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSer---AspIle 191  
Db 621 AAGTTGCCAAAGTAGTAACCCAGNAATCTGGAGGTGCAGGCGGAGCGGGGTGTG 680  
Qy 192 LysLeuGluLysProGlnGluGlnThrValSerGlu-----GlyGly 205  
Db 681 GATCTCCATGAACAGTCCAGCAAGAACAGAGCCAGTGTGTAGCAACTAGGGGGGTGC 740  
Qy 206 -----Cys 206  
Db 741 TAGCAGCAACAAAGTATGAGCTAGCACAGAGCTAAGAAATAACCTCCATCCCTACCCC 800

```
Qy 207 SerCysTyrSerProMetSerSerThrLeuProGln-----LysProPro 222
Db 801 TCAGCAACAACCCCTACGGTAACAGCACATGACCCCTGGCTCCCAAGGCGCTGCTCTCT 860
Qy 223 TyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeuIleThr 242
Db 861 GACAGCTCGGTCATGCGACTTTTAAACGCTTCAGCAACAACACAGGCGAGCTGTGCA 920
Qy 243 PheCysAsnSerSerLeuLeu 249
Db 921 CTGGCTCTCTACCCCTACTC 941

RESULT 9
US-09-949-016-409
; Sequence 409, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 409
; LENGTH: 1630
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-409

Alignment Scores:
Pred. No.: 4,75e-37 Length: 1630
Score: 371.50 Matches: 77
Percent Similarity: 58.33% Conservative: 35
Best Local Similarity: 40.10% Mismatches: 79
Query Match: 28.49% Indels: 1
DB: 4 Gaps: 1

US-09-441-857-12 (1-254) x US-09-949-016-409 (1-1630)

Qy 13 LysPheLysLeuValPheLeuGlyGluGlnSerValAlaLysThrSerLeuIleThrArg 32
Db 78 CAGTTCAAAATGGTCTCTGGGAGAACTCTGAGTGGGAAAGTCAAGCTGTGTATACGT 137
Qy 33 PheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyIleAspPheLeuSer 52
Db 138 TTTGTCAAGGCGACTTCATCAGTACAGGAGACCATTTGGAGCGCTTCTTACC 197
Qy 53 LysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrAspThrAlaGlyGln 72
Db 198 CAGTCCGTTTGTCTAGATGACACACAGTGAAGTTTGAGATCTGGGACACAGCTGGGCGAG 257
Qy 73 GluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaValValVal 92
Db 258 GAGCGATATACAGCTTAGCCCTCATGTACTACAGGGGTGCCCAAGCTGCAATCGTGT 317
Qy 93 TyrAspIleThrAsnValAsnSerPheGlnGlnThrThrLysTrpIleAspAspValArg 112
Db 318 TACGACATTACTAATCAGAAACCTTTGCCGAGCAAGACATGGTGAAGAACTACAG 377
Qy 113 ThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeuAlaAsp 132
Db 378 CGACAGGCGAGCTCCTAGCATCGTTATTGCTCCCTGGCAGGGAACAAAGCTGACCTGGCCAAC 437
Qy 133 LysArgGlnValSerValGluGluGlyAlaLysAlaLysGlyLeuAsnValThrPhe 152
```

```
Db 438 AAACGTATGGTGGAGTATGAGAGCCGAGCATATGAGATGACACAGCTTATTGTTTC 497
Qy 153 IleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgValAlaAla 172
Db 498 ATGAGAGACTTCAGCCAAAGACAGCTATGAACGTGAATGATCTCTTCTGGCAATAGCTAAG 557
Qy 173 AlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSer--AspIle 191
Db 558 AAGTTGCCAAAGAGTGAACCCGAGAAATCTGGAGGTGCAGCGCCGAGCCGCGGTGTG 617
Qy 192 LysLeuGluLysProGlnGluGlnThrValSerGlu 203
Db 618 GATCTCCATGACAGTCCGAGCAGAACAGAGCCAG 653
```

```
RESULT 10
US-09-016-434-1124
; Sequence 1124, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA: US/09/016,434
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1124:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: GI388194
US-09-016-434-1124
```

```
Alignment Scores:
Pred. No.: 2,81e-35 Length: 921
Score: 354.50 Matches: 67
Percent Similarity: 60.32% Conservative: 47
Best Local Similarity: 35.45% Mismatches: 72
Query Match: 27.19% Indels: 3
DB: 4 Gaps: 1

US-09-441-857-12 (1-254) x US-09-016-434-1124 (1-921)
```



QY 11 LeuArgLysPheLysLeuValPheLeuGlyGluInSerValAlaLysThrSerLeuLeu 30  
Db ATACGGAGCTCAAAAGTGGCTTCTCGGGACACATGGGGTGGGAAATCAAGCATCGG 126  
QY 31 ThrArgPheArgTyrAspSerPheAspThrTyrGlnAlaIleGlyLeuAspPhe 50  
Db TGTCGATTGTCAGGATCACATTTGACCAACAACATCAGCCCTACTATTGGGGCATCTTT 186  
QY 51 LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTrpAspThrAla 70  
Db ATGACCAAAACGTGCGCTTGTGGAATGAACCTTCAACAGTTCTCATCTGGGACACTGCT 246  
QY 71 GlyGlnGluArgLeuSerLeuLeuProArgTyrIleArgAspSerAlaAlaVal 90  
Db GGTCAAGAACGGTTTCATTGCTCCCATGCTACTATCGAGGCTCAGCTCAGCTGTT 306  
QY 91 ValValTyrAspIleThrAsnValAsnSerPheGlnInThrThrLysTrpIleAspAsp 110  
Db ATCGGTATGATATTACCAAGCAGGATTCATTTTATACCTTGAAGAAATGGGTCAAGGAG 366  
QY 111 ValArgThrGluArgGlySerAspValIleThrLeuValGlyAsnArgThrAspLeu 130  
Db CTGAAGAACATCGTCCAGAAACATTTGTAATGGCCATCGCTGGAAACAAGTGCACCTC 426  
QY 131 AlaAspLysArgGlnValSerValGluGluGlyGluArgLysAlaLysGlyLeuAsnVal 150  
Db TCAGATATTAGGAGGTTCCCTGGAAGGATGCTAAGGAATACGCTGAATCCATAGTGCC 486  
QY 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170  
Db ATCGGTGTGACAGCAAGTGCAAAAAATGCTATTATATATCGAAGAGCTCTTTCAAGGAATC 546  
QY 171 AlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190  
Db AGCCCCAGATCCACCTTGAGCCCAATGAAATGAAACAAT-----GGAACA 597  
QY 191 IleLysLeuGluLysProGlnGluGln 199  
Db ATCAAAGTTGAGAGCAACCAACCATGCAA 624

## RESULT 11

US-09-949-016-4216  
; Sequence 4216, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4216  
; LENGTH: 773  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-4216

Alignment Scores:  
Pred. No.: 1.66e-34 Length: 773  
Score: 347.50 Matches: 79  
Percent Similarity: 50.67% Conservative: 35  
Best Local Similarity: 25.11% Mismatches: 86  
Query Match: 26.65% Indels: 25  
DB: 4 Gaps: 4

US-09-441-857-12 (1-254) x US-09-949-016-4216 (1-773)  
QY 14 PheLysLeuValPheLeuGlyGluInSerValAlaLysThrSerLeuIleThrArgPhe 33  
Db TTCAAGATTCATCATCGGACAGCAGCGTGGCAGAGCTCTTCTCTTCCGCTAT 161  
QY 34 ArgTyrAspSerPheAspAsnThrTyrGlnAlaIleGlyLeuAspPheLeuSerLys 53  
Db GCTCAGCACTCGTTCAAGCTTCTGTCAGCACCGTGGCATCGACTTCAAGGTCAAG 221  
QY 54 ThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGlu 73  
Db ACCATCTATCCCAACGACAGAGGATCAAGCTCGAGATCTCGGACACACGAGGCAAG 281  
QY 74 ArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaValValValTyr 93  
Db CGGTACCGGACCATCACCGCATCTACCGGGCGCTATGGGCTTCTCATGTAT 341  
QY 94 AspIleThrAsnValAsnSerPheGlnInThrThrLysTrpIleAspAspValArgThr 113  
Db GACATCAACCAAGGAAATCTTCAATGCAGTCAGGACTGCTCCACCAGATCAAGACC 401  
QY 114 GluArgGlySerAspValIleThrLeuValGlyAsnArgThrAspLeuAlaAspLys 133  
Db TACTCATGGGCAATGCCAGTCTGCTGGTAGGAAACAAAGTGTGACATGGAGGATGAG 461  
QY 134 ArgGlnValSerValGluGlyGluArgLysAlaLysGlyLeuAsnValThrPheIle 153  
Db CGGTGTGTCTCATCAGAACGTGGCGGAGTAGCTGACACCTTGGGTTCGAGTCTTT 521  
QY 154 GluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgValAla----- 171  
Db GAGCAAGCGCAAGGACCAACATTAACTGTCAGCAGACCTTTGAGCGCTGCTGGATGTC 581  
QY 172 -----AlaAlaLeuProGlyMetGluSerThrGln 181  
Db ATCTCGGAGAAAGATGTCGAGTCGTGGACACGGCGGACCTGCGGTCCACAGGCCCAAG 641  
QY 182 AspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThrVal 201  
Db CAGGCGCCACAG-----CTCAGTGACCAAGGAGTGCCACCGCACCAGGAC----- 686  
QY 202 SerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThr-----Leu 217  
Db -----TGGCCTGCTGAGAGCAATCCCACTCCCTTTCCCTCTTCCCTGCTCTT 734  
QY 218 ProGlnLysProPro 222  
Db CCCACCTTCCCCCA 749

## RESULT 12

US-09-248-796A-5859  
; Sequence 5859, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 5859  
; LENGTH: 717  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-5859

Alignment Scores:



```
/ APPLICANT: Yang, Yonghong
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yuning
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: John Tillinghast
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: No. 659662el Nucleic Acids and
/ POLYPEPTIDES
/ FILE REFERENCE: 784CIP2B
/ CURRENT APPLICATION NUMBER: US/09/620,312D
/ PRIORITY FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIORITY FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
/ PRIORITY FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 1105
/ SOFTWARE: pt_FL_genes Version 1.0
/ SEQ ID NO 587
/ LENGTH: 1631
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (66)..(650)
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)....(1631)
/ OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-587

Alignment Scores:
Pred. No.: 2,42e-33 Length: 1631
Score: 342.50 Matches: 80
Percent Similarity: 49.43% Conservative: 50
Best Local Similarity: 30.42% Mismatches: 108
Query Match: 26.27% Indels: 25
DB: 4 Gaps: 3

US-09-441-857-12 (1-254) x US-09-620-312D-587 (1-1631)

Qy 3 AlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGlyGluIn 22
Db 48 GCCCTCGGTCCTCCGGGCCATGCGCTGAGGAGCTCAAAGTGTCTGCTCGGGGATACA 107
Qy 23 SerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyr 42
Db 108 GGTGTAGGTAAATCGAGTATTGTGTGGCGGTTTGTGGAGACAGTCTTTGATCCAAACATC 167
Qy 43 GlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIle 62
Db 168 AACCCAAACATAGGGGCATCTTTATGACCAAGCTGCTCCAGTACCAAAATGAGGTACAT 227
Qy 63 GlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLysLeuIleProArgTyr 82
Db 228 AAATCTCTAATCTGGGATACAGCTGGCAAGAACGATTTCGTGCTTAGCACCATTGTAC 287
Qy 83 IleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSerPheGln 102
Db 288 TATCGAGGTCGGCTGCGACTATAATCGTTTATGATATCACAAAAGAGAGACATTTTCA 347
Qy 103 GlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIleIleThr 122
Db 348 ACATTAAAGAATTGGGTGAAGAGCTTCCAGCATGCGCCACCTTAATATTGTAGTGGC 407
Qy 123 LeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlyGlu 142
Db 408 ATTGAGGAATAAATGTGATCTTATCGATGTAAGAGAAGTCATCGAGAGAGATCCAAAG 467
Qy 143 ArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGlyTyrAsn 162
Db 468 GACTACGCCGACTCTATTATTCGAATTTTGTAGAGACAGCGCAAAAAACGCGATAAAC 527

163 VallysGlnLeuPheArgValAlaAlaAlaLeuPro----- 175
528 ATAAATGAACCTCTTTATAGAAATTAGTCGAAGAATTCATCCACTGACGCCAACCTGCCA 587
176 -----Gly-MetGluSerThrGlnAspGlySerArgGluAspMetSer-AspIleLysL 193
588 TCTGCGCGTAAGGGCTTCAAACTCCGAAGACAGCCTTCAGAGCCAAAGCGAGCTGCTGC 647
193 euGluLysProGlnGluGlnThrValSerGlu----- 203
648 TGACCGAACCTCAGCCTCTCAGACTTGATGATGAAGTAGGTAGGTGCTCCTGAAAGTTAACAGG 707
204 -----GlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLysP 221
708 AGGGCTGGGGTCCCTGCCACCACTTTTTCACCTAGCCAGCTCTTGAGTCTTCTCCGGTCAAA 767
221 roProTyrSerPheIleAsp-----CysSerValAsnIleGlyLeuAsnLeuPheProS 239
768 AAGGATTCACAGAAATGACCACTTCTGTTCTCCAAAGACTGCAGCAATGATATTTCAGT 827
239 erLeu 240
828 CTGTG 832

RESULT 15
US-09-270-767-1882/c
; Sequence 1882, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1882
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1882

Alignment Scores:
Pred. No.: 4.12e-33 Length: 998
Score: 338.00 Matches: 73
Percent Similarity: 57.64% Conservative: 44
Best Local Similarity: 35.96% Mismatches: 82
Query Match: 25.92% Indels: 4
DB: 4 Gaps: 2

US-09-441-857-12 (1-254) x US-09-270-767-1882 (1-998)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 883 ATGAGCTCGACAGAACGAGGAGCGCCCGCCAGCTTAATTTCAAGCGCGTGTCTGCTGGCG 824
Qy 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40
Db 823 GAAAGTTGTGGGCGACAGACGCTCGTGTGCTGCGCTACATGGAAGACCGGTTCATATGCC 764
Qy 41 ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 763 CAGCACCTAACGACCCCTCGAGCTTCTTTGTCAGCCGCAAGATGTCCTCGAGGATGGG 704
Qy 61 ---ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuAspSerLeuIle 79
Db 703 AGAAGGCGCAGTGTGAATATTGGGACACGCGTGTGTCAGGAGCGGTTCCACGCCCTGGGA 644
Qy 80 ProArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsn 99
Db 643 CCCATCTACTATCGCGCTCCGAGCGCCCTGCTGCTATGATACATACGACCGACGAGAC 584
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 25, 2005, 04:55:10 ; Search time 6740 Seconds  
(without alignments)  
228.790 Million cell updates/sec

Title: US-09-441-857-12  
Perfect score: 1304  
Sequence: 1 MSAGDGFNPLKPKLVFLG.....NLPLSLITFCNSLLPVSWR 254

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_epool\_p/US09441857/runat\_22042005\_155511\_24863/app\_query.fasta\_1.391  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MTNMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -NATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODS=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTNLEN=0  
-MAXLEN=2000000000 -USER=US09441857\_@CGN 1 1 723\_@runat\_22042005\_155511\_24863  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
C 1	99.6	1299	744802	17	US-10-292-798-1369
2	1133	86.9	3118	9	US-09-925-300-424
3	959	73.5	740	17	US-10-116-275-157
4	892	68.4	1266	17	US-10-172-118-1656
5	892	68.4	1266	17	US-10-342-887-1656
6	892	68.4	1266	19	US-10-848-755A-181
7	846.5	64.9	1157	13	US-10-108-605-44
8	835.5	64.1	2456	17	US-10-094-749-430
9	715	54.8	926	18	US-10-653-047-7008
C 11	709.5	54.4	1175	17	US-10-437-963-63907
12	709.5	54.4	1200	17	US-10-425-114-4106
13	709.5	54.4	1220	17	US-10-425-114-1068
14	707	54.2	1242	18	US-10-425-115-20462
15	705.5	54.1	1308	17	US-10-425-114-27873
16	704.5	54.0	1145	18	US-10-425-115-160784
17	699.5	53.6	1069	17	US-10-425-114-27069
18	697	53.5	1317	17	US-10-424-599-47043
19	688	52.8	1155	17	US-10-424-599-47044
20	667	51.2	523	18	US-10-021-323-14369
21	662.5	50.8	714	17	US-10-424-599-47503
C 22	653	50.1	1045	18	US-10-437-963-43192
23	649.5	49.8	1040	17	US-10-425-114-28361
24	649.5	49.8	1116	18	US-10-425-115-181339
25	642	49.2	443	9	US-09-960-352-7355
26	623	47.8	1296	18	US-10-739-930-4235
27	596	45.7	791	18	US-10-767-795-6480
28	593.5	45.5	3826	15	US-10-017-161-1727
29	593.5	45.5	3826	17	US-10-292-798-1383
30	582	44.6	648	17	US-10-369-493-25486
31	572	43.9	846	17	US-10-369-493-25812
32	569	43.6	424	10	US-09-918-995-5540
33	562	43.1	714	18	US-10-653-047-213
34	541.5	41.5	574	18	US-10-767-795-6479
35	536.5	41.1	5158	17	US-10-275-933-11
36	532	40.8	1072	17	US-10-424-599-72715
37	485	37.2	570	18	US-10-437-963-43190
38	467	35.8	525	17	US-10-369-493-36580
39	461.5	35.4	963	17	US-10-424-599-47042
C 40	431	33.1	320	9	US-09-867-701-8545
41	431	33.1	1042	17	US-10-260-238-3339
42	396	30.4	819	17	US-10-262-511-221
43	396	30.4	960	17	US-10-262-511-217
44	396	30.4	1007	9	US-09-822-849A-563
45	396	30.4	1529	17	US-10-205-331-71

ALIGNMENTS

RESULT 1

US-10-292-798-1369/c  
; Sequence 1369, Application US/10292798  
; Publication No. US20030235833A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1369



QY 201 ValSerGluGlyCysSerCysTyrSerProMetSerSerThrLeuProGlnLys 220  
Db 345108 GTCCAGCAAGGGGTTGTTCCCTGCTACTCTCCCAATGTCATCTTCAACCCCTCCTCAGAAG 345049  
QY 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240  
Db 345048 CCCCCTTACTCTTTTCATTGACTGCGAGTGTGAATATTGGCTTGAACCTTTTTCCTCTCATTTA 344989  
QY 241 IleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254  
Db 344988 ATAACGTTTGTCAATTCATCTTGCCTGCTGCTCGTGGAGG 344947

## RESULT 2

US-09-925-300-424  
; Sequence 424, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 424  
; LENGTH: 3118  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (388)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (485)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-424

## Alignment Scores:

Pred. No.: 7,3e-134 Length: 3118  
Score: 1133.00 Matches: 224  
Percent Similarity: 89.53% Conservative: 7  
Best Local Similarity: 86.82% Mismatches: 23  
Query Match: 86.89% Indels: 4  
DB: 9 Gaps: 1

US-09-441-857-12 (1-254) x US-09-925-300-424 (1-3118)

QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 415 ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 474  
QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40  
Db 475 GAGCAAGCAGTGGGAAGACATCTTGTGATCACCAGATTTCATGTATGACAGTTTTCACAC 534  
QY 41 ThrTyrGlnAlaIleIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 535 ACCTATCAGGCAACAATTTGGCATTGACTTTTATCAAAAATATGTACTTGGAGGATCGA 594  
QY 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
Db 595 ACAGTACGATTGCAATTATGGGACACAGCAGGTTCAGAGCGGTTCAGGAGCTTGATTCT 654  
QY 81 ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer 100  
Db 655 ACCTACATTCTGATCTCCACTGTGGCAGTTGTTGTTATGATATCAAAATGTTAACTCA 714  
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120

Db 715 TTCCAGCAAACTCAAAAGTGGATTGATGATGTCAAGACAGAAAGAGGAGTGAATGATTTATC 774  
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysAlaGlnValSerValGluGlu 140  
Db 775 ATCATGCTAGTAGGAATACAGATCTTGCTGACAGAGGCAAGTGTCAATTGAGGAG 834  
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 835 GGAGAGAGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTGTCAAAAGCTGGA 894  
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180  
Db 895 TACAATGTAAAGCAGCTCTTTTCGACGTGTACAGCAGCTTTGCCGGGAATGGAAGACACA 954  
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200  
Db 955 CAGGACAGACGAGAGAAGATATGATTGACATAAACTGGAAAGCCTCAGGAGGACCA 1014  
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220  
Db 1015 GTCAGTGAAGGAGGCTGTTCTGCTTAATCTCCCATGTCTCTTCAACCTTCTTCAAGAGC 1074  
QY 221 -----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236  
Db 1075 TCACCTGCTTTGGCCCCCTTACTCTTTCATTGACTGCAGTGTGAATATTGGCTTGAACCTT 1134  
QY 237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254  
Db 1135 TTCCCTTCAGTAATAACGTAATTGCAATTCATCTTGCCTGCTGCTCGTGAGA 1188

## RESULT 3

US-10-116-275-157  
; Sequence 157, Application US/10116275  
; Publication No. US20030211476A1  
; GENERAL INFORMATION:  
; APPLICANT: Elan Pharmaceutical Technology  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Brayden, David  
; APPLICANT: Byrne, Daragh  
; APPLICANT: Lambkin, Imelda  
; APPLICANT: Higgins, Lisa  
; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and  
; FILE REFERENCE: E1067/20087  
; CURRENT APPLICATION NUMBER: US/10/116,275  
; CURRENT FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 157  
; LENGTH: 740  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-116-275-157

## Alignment Scores:

Pred. No.: 1.62e-112 Length: 740  
Score: 959.00 Matches: 191  
Percent Similarity: 90.28% Conservative: 4  
Best Local Similarity: 88.43% Mismatches: 21  
Query Match: 73.54% Indels: 0  
DB: 17 Gaps: 0

US-09-441-857-12 (1-254) x US-10-116-275-157 (1-740)

QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 71 ATGTCCACGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 130  
QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40  
Db 131 GAGCAAGCAGTGGGAAGACATCTTTGATCACCAGATTTCATGATGATGACAGTTTTCACAC 190

```
QY 41 ThrTyrGlnAlaIlelleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 191 ACCTATCAGGCAACAAATTTGGCATTGACCTTTTATCAAAACCTAATGACTTGGAGATCGA 250
QY 61 ThrIleGlyLeuArgLeuThrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 251 ACAGTACGATTTCAATATTATGGGACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTGATTCT 310
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 311 AGCTACATTCGAGACTCCACTGTGGCAGTGTGTGTATGATATCACAAATGTTAACTCA 370
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 371 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAGAAACAGAGGAGGATGATGTTATC 430
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 431 ATCATGCTAGTAGGAATAAAACAGATCTTGCTGACAAAGAGCAAGTGTCAATTGAGGAG 490
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 491 GGAGAGAGGAAGCCAAAGAGCTGAATGTTATGTTATTTGAAACTAGTGCAAAAGCTCGA 550
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180
Db 551 TACATGTTAAGCAGCTCTTTCGACGTGTAGCAGCAGCTTTGCCGGGAATGGAAGCACA 610
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 611 CAGGACAGAAGCAGAGAATATGATTGACATATAAACTGGAAAAAGCCTCAGGAGCAACA 670
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThr 216
Db 671 CTCAGTGAAGGAGGCTGTCTCTGCTTAATGTCCTAGTATCTTCAACC 718
```

## RESULT 4

```
US-10-172-118-1656
; Sequence 1656, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1656
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 016577
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1656
```

```
Alignment Scores:
Pred. No.: 1,45e-103 Length: 1266
Scores: 892.00 Matches: 172
Percent Similarity: 90.38% Conservative: 16
Best local Similarity: 82.69% Mismatches: 20
Query Match: 68.40% Indels: 0
Db: 17 Gaps: 0
```

```
US-09-441-857-12 (1-254) x US-10-172-118-1656 (1-1266)
QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 338 ATGTCGCCAGGGGAGATTTTGGGAATCCACTGAGAAAAATTCAGAGTTGGTGTCTTCTGGGG 397
QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40
Db 398 GAGCAGAGCGTCGGGAGAGCGTCTCTGATTACGAGGTTTCATGTACGACAGCTTCGACAC 457
QY 41 ThrTyrGlnAlaIlelleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 458 ACATACCAGGCAACCATTTGGGATTGACTTCTTGTCAAAAACCATATCTTGGAGACCGC 517
QY 61 ThrIleGlyLeuArgLeuThrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
Db 518 ACGTGGCAGCTGCAGCTCTGGGACACAGCTGCTCAGGAGAGGTTCGCGAGCTCATCCCC 577
QY 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
Db 578 AGCTACATCCGGGACTCCACGGTGGCTGTGGTGTGTACGACATCACAAATCTCACTCC 637
QY 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 638 TTCCAACAGACCTCTAAGTGGATCGACGACGTCAGGACAGAGGGGCGAGTGTATTATC 697
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 698 ATCATGCTGTGGGCCAACAGACGACCTGGCTGATTAAGAGGCAGATAACCATCGAGGAG 757
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 758 GGGAGCAGCGCCGCCAAAGAACTGAGCGTCTATGTTATTGAGACCACTGCGAAGACTGGC 817
QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180
Db 818 TACAACGTGAAGCAGCTTTTTCGACGCTGTGGCGTGGCTCTACCCGGAATGGAGAAATGC 877
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 878 CAGGAGAAAAACAAAGAGGATGATCGACATCAAGCTGGACAAACCCAGGAGCCCCCG 937
QY 201 ValSerGluGlyGlyCysSerCys 208
Db 938 GCCAGCGAGGCGGCTGCTCTCTCTGC 961
```

## RESULT 5

```
US-10-342-887-1656
; Sequence 1656, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1656
; LENGTH: 1266
; TYPE: DNA
```







## RESULT 9

US-09-441-857-12 (1-254) x US-10-653-047-7008 (1-926)

Alignment Scores:	2,81e-80	Length:	1175
Pred. No.:	710.00	Matches:	139
Score:	75.59%	Conservative:	26
Percent Similarity:	63.76%	Mismatches:	47
Best Local Similarity:	54.45%	Indels:	2
Query Match:	18	Gaps:	6
DB:			
US-09-441-857-12 (1-254) x US-10-437-963-63907 (1-1175)			

```
QY      11  LeuArgLysPheLysLeuValPheLeuGlyGluGlnSerValAlaLysThrSerLeuIle 30
      1065  CTCGCCAGTACAGCTGGTGTCTCTCGCGCACCAGTCGTCGCGCAAGCAGCATATC 1006
QY      31  ThrArgPheArgTyrAspSerPheAspAenThrTyrGlnAlaIleIleGlyLeuAspPhe 50
      1005  ACCCGCTTCATGTACGACAAGTTCGACAACACCTACAGGCTACAAATCGGTATTGACTTC 946
QY      51  LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrAspThrAla 70
      945  CTGTCAAAAGACAATGTACCTCAAGATAGGACTGTACAGACTGCAACTTTGGGATACAGCT 886
QY      71  GlyGlnGluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaVal 90
      885  GGTCAAGAACGATTCAGAGATTGATCCAGCTATATCAGAGACTCTTCGGTTGCTGTC 826
QY      91  ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrLysTyrIleAspAsp 110
      825  ATCGTATTGATGTTGCAAGCAGACAGCTTTCTTAACACTTCAAAATGGATAGAGAA 766
QY      111  ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
      765  GTTAGAACTAGAGGGGCGAGTGATCTTATCATTTGTCTGGTTGGGAACAAAACTGATCTT 706
QY      131  AlaAspLysArgGlnValSerValGluGlyGluArgLysAlaLysGlyLeuAsnVal 150
      705  GTTGCAAGAGGCAAGTCTCCATAGAGGAGGAGGCAAGCGAAAGATCTCGGTGTG 646
QY      151  ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
      645  ATGTTTCATTTGAACACAGTCTTAAGCTGGTTTAACTAATAGGCGTTGTTCCGTAAGATT 586
QY      171  AlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
      585  CGCGTGTCACCTTCTCGGAATGAGACCCCTCTCATCAGCAAGACGAGGACATGGTTGAT 526
QY      191  IleLysLeuGlu-----LysProGlnGluGlnThrValSerGluGlyGlyCys 206
      525  GTGAACCTTGAAGTCAAGCAAGCCCACTCATCCAGTCTCAGGCACAGGCTGGGGGATGC 466
QY      207  SerCys-----TyrSerProMetSerSerThrLeuProGlnLysProPro 222
      465  AGTTGTTAGTTCAGTCCCTTAACTTACTGCTTCGCGGTCTCAAGGCCCCCC 412
```

## RESULT 11

```
US-10-425-114-4106
; Sequence 4106, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 4106
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700349145_FLI
US-10-425-114-4106
```

```
Alignment Scores:      3.36e-80      Length:      1200
Pred. No.:      709.50      Matches:      144
Score:
```

```
Percent Similarity:      72.46%      Conservative: 27
Best Local Similarity: 61.02%      Mismatches: 50
Query Match:      54.41%      Indels: 15
DB:      17      Gaps: 3
US-09-441-857-12 (1-254) x US-10-425-114-4106 (1-1200)
QY      11  LeuArgLysPheLysLeuValPheLeuGlyGluGlnSerValAlaLysThrSerLeuIle 30
      181  CTCGCCAGTACAGCTGGTGTCTCTCGCGGACAGTCCGTCGCGCAAGCAGCATATC 240
QY      31  ThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLeuAspPhe 50
      241  ACCCGCTTCATGTACGACAAGTTCGATACACCTTACAGGCTACGATTTGATTGATTTC 300
QY      51  LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrAspThrAla 70
      301  CTGTCAAAAGACAATGTACCTTGAAGATAGAATCTGTGAGACTTCCAACCTCGGATACCTGCT 360
QY      71  GlyGlnGluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaVal 90
      361  GGTCAAGAAAGATTTCAGAGTTTAAATTCGAAGCTATATTAGAGACTCTTCAGTTGCTGTC 420
QY      91  ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrLysTyrIleAspAsp 110
      421  ATGTATTTCGATGTTGCAAGCAGCAGCTCTCTTAATAATACATCTAAGTGGATAGAGAA 480
QY      111  ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
      481  GTTCGCACCTGAGAGGGGCGAGTGATCTTATCATTTGTCTTGGGAACAAAACTGACCTT 540
QY      131  AlaAspLysArgGlnValSerValGluGlyGluArgLysAlaLysGlyLeuAsnVal 150
      541  GTTGACAAGAGGCAAGTCTCAATAGAGAAAGGGGAGGCAAGCGCAAGGACCTTTGGAGTC 600
QY      151  ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
      601  ATGTTTATTGAACACAGTGTCTAAGCTGGGTTTAACTAATAGGCGCTGTTCCGTAAATTT 660
QY      171  AlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
      661  GCTGTCGACTTCTCGAATGGAGACGCTCTCATCAGCAAGCAGGAAAGACATGGTTGAT 720
QY      191  IleLysLeuGlu-----LysProGlnGluGlnThrValSerGluGlyGlyCys 206
      721  GTGAACCTGAGGTCGCGCAATGCAAACTCGTCCAGTCTCAGGCTCAGGCTGGGGGATGC 780
QY      207  SerCysTyrSerProMetSerSerThrLeuProGlnLysProProTyrSerPheIle 226
      781  AGTTGT-----TAGTTGCGAGTCTCTGACACTTGTCT----- 810
QY      227  AspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeuIleThr 242
      811  ---TGCTCATGATCTTTGGGACTCTGAGCATCAGAGGCGGTATCACT 855
```

## RESULT 12

```
US-10-425-114-1068
; Sequence 1068, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1068
```

```

; LENGTH: 1220
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700101038_FLI
US-10-425-114-1068

Alignment Scores:
Pred. No.: 3,45e-80 Length: 1220
Score: 709.50 Matches: 144
Percent Similarity: 72.46% Conservative: 27
Best Local Similarity: 61.02% Mismatches: 50
Query Match: 54.41% Indels: 15
DB: 17 Gaps: 3

US-09-441-857-12 (1-254) x US-10-425-114-1068 (1-1220)

QY 11 LeuArgLysPheLysLeuValPheLeuGluGlnSerValAlaLysThrSerLeuLeu 30
DB 197 CTCGCCAAGTACAAGCTCGTCTCTCTGGGGGACCAAGTCCGTCGCGCAAGACGAGCATCATC 256

QY 31 ThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLeuPhe 50
DB 257 ACCCGCTTCATGTACGACAAAGTTCGATTAACCTTACCAGGCTACGATTGGTATTGATTTC 316

QY 51 LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrPheThrAla 70
DB 317 CTGTCAAGACAAATGTACCTTCAAGATAGAACTGTGAGACTCCAACTCTGGGATACTGCT 376

QY 71 GlyGlnGluArgLeuArgSerLeuLeuProArgTyrIleArgAspSerAlaAlaVal 90
DB 377 GGTGAGAAAGATTACAGGAGTTAAATTCGAAGCTATATTAGAGACTCTTCAGTTGCTGTC 436

QY 91 ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrThrLysThrIleAspAsp 110
DB 437 ATTGATTTCGATGTTGCAAGCAGGAGCTCTTCTTAAATACATCTAAAGTGGATAGAGAA 496

QY 111 ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
DB 497 GTTCGACTGAGAGGGGCGAGTGATGTTATCATTTGCTTGTGGGAACAACATGACCTT 556

QY 131 AlaAspLysArgGlnValSerValGluGluGlyLysAlaLysGlyLeuAsnVal 150
DB 557 GTTGACAAGAGCAAGTCTCAATAGAGGAAGGGGAGGCAAGCGCAAGACCTTCGAGTC 616

QY 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
DB 617 ATGTTTATTGAACCAAGTCTAAGGCTGGGTTTAACTTAAGCGCTGTTCCGTAAATTT 676

QY 171 AlaAlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
DB 677 GCTGCTGCACCTTCCTGGAATGAGAGCGCTCTCATCAGCGAAGCAGGAGACATGTTGAT 736

QY 191 IleLysLeuGlu-----LysProGlnGluGlnThrValSerGluGlyCys 206
DB 737 GTGAACCTTGAAGTCCGCGCAATGCAAACTCGTCCAGTCTCAGGCTCAGGCTGGGGGATGC 796

QY 207 SerCysTyrSerProMetSerSerThrLeuProGlnLysProProTyrSerPheIle 226
DB 797 AGTTGT-----TAGTTGCAGTCTCTGACACTTGCT----- 826

QY 227 AspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeuIleThr 242
DB 827 ---TGCTCATGATCTTTGGGACTCTGAGCATCAGAAGGCGTTACTACT 871

RESULT 13
US-10-425-115-20462
; Sequence 20462, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 04-28
; SEQ ID NO 20462
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1506)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_118663C.1
US-10-425-115-20462

Alignment Scores:
Pred. No.: 4.77e-80 Length: 1506
Score: 709.50 Matches: 144
Percent Similarity: 72.46% Conservative: 27
Best Local Similarity: 61.02% Mismatches: 50
Query Match: 54.41% Indels: 15
DB: 18 Gaps: 3

US-09-441-857-12 (1-254) x US-10-425-115-20462 (1-1506)

QY 11 LeuArgLysPheLysLeuValPheLeuGluGlnSerValAlaLysThrSerLeuLeu 30
DB 335 CTCGCCAAGTACAAGCTCGTCTCTCTGGGGGACCAAGTCCGTCGCGCAAGACGAGCATCATC 394

QY 31 ThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLeuPhe 50
DB 395 ACCCGCTTCATGTACGACAAAGTTCGATTAACCACTTACCAGGCTACGATTGGTATTGATTTC 454

QY 51 LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrPheThrAla 70
DB 455 CTGTCAAGACAAATGTACCTTGAAGATAGAACTGTGAGACTCCAACTCTGGGATACTGCT 514

QY 71 GlyGlnGluArgLeuArgSerLeuLeuProArgTyrIleArgAspSerAlaAlaVal 90
DB 515 GGTGAGAAAGATTACAGGAGTTAAATTCGAAGCTATATTAGAGACTCTTCAGTTGCTGTC 574

QY 91 ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrThrLysThrIleAspAsp 110
DB 575 ATTGATTTCGATGTTGCAAGCAGGAGCTCTTCTTAAATACATCTTAAGTGGATAGAGAA 634

QY 111 ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
DB 635 GTTCGACTGAGAGGGGCGAGTGATGTTATCATTTGCTTGTGGGAACAACATGACCTT 694

QY 131 AlaAspLysArgGlnValSerValGluGluGlyLysAlaLysGlyLeuAsnVal 150
DB 695 GTTGACAAGAGCAAGTCTCAATAGAGGAAGGGGAGGCAAGCGCAAGACCTTCGAGTC 754

QY 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
DB 755 ATGTTTATTGAACCAAGTCTAAGGCTGGGTTTAACTTAAGCGCTGTTCCGTAAATTT 814

QY 171 AlaAlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
DB 815 GCTGCTGCACCTTCCTGGAATGAGAGCGCTCTCATCAGCGAAGCAGGAGACATGTTGAT 874

QY 191 IleLysLeuGlu-----LysProGlnGluGlnThrValSerGluGlyCys 206
DB 875 GTGAACCTTGAAGTCCGCGCAATGCAAACTCGTCCAGTCTCAGGCTCAGGCTGGGGGATGC 934

QY 207 SerCysTyrSerProMetSerSerThrLeuProGlnLysProProTyrSerPheIle 226
DB 935 AGTTGT-----TAGTTGCAGTCTCTGACACTTGCT----- 964
```

```
Qy 227 AspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeuIleThr 242
Db 965 ---TGCTCATGATCTTTGGGACTCTGAGCATCAGAGCGGTATCACT 1009

RESULT 14
US-10-425-115-20463
; Sequence 20463, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 20463
; LENGTH: 1242
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_118664C.1
US-10-425-115-20463

Alignment Scores:
Pred. No.: 7,41e-80 Length: 1242
Score: 707.00 Matches: 136
Percent Similarity: 78.22% Conservative: 22
Best Local Similarity: 67.33% Mismatches: 40
Query Match: 54.22% Indels: 4
DB: 18 Gaps: 1

US-09-441-857-12 (1-254) x US-10-425-115-20463 (1-1242)

Qy 11 LeuArgLysPheLysLeuValPheLeuGlyGlnSerValAlaLysThrSerLeuIle 30
Db 193 CTCGCCAAGTACAAAGCTCGTCTTCTGGGGAGCAGCTCCGCGGCAAGCAGCATCATC 252

Qy 31 ThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLeuAspPhe 50
Db 253 ACCCGCTTCATGATAGTAAAGTTCGACAACTTACCGAGGTACAAATTTGGTATTGATTTC 312

Qy 51 LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrPheAspThrAla 70
Db 313 CTGTCAAAAGACAATGTACTTGAAGATAGAACTGTGAGACTCCAACCTCTGGGATACAGCT 372

Qy 71 GlyGlnGluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaVal 90
Db 373 GGTGAGGAAAGGTTTCAGGAGTTAAATTCCAAGCTATATACAGAGACTCTTCAGTTGCTGTC 432

Qy 91 ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrLysThrLysTrpIleAspAsp 110
Db 433 ATTGTATTCGATGTTCGAGCAGGAGCTTTCTTAATATCTCCAGTGTGATAGATGAA 492

Qy 111 ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
Db 493 GTTCGCACTGAGAGGGGCGAGTGATTTATCTTGTCTTGTGGGAACAAACTGACCTT 552

Qy 131 AlaAspLysArgGlnValSerValGluGlyGlnArgLysAlaLysGlyLeuAsnVal 150
Db 553 GTTGCAAGAGCGCAAGTCTCGATAGAGGAAGGAGGCAAGGCAAGGACCTTTGGTGTGTC 612

Qy 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
Db 613 ATGTTTATTGAACACAGTGTCTTAAGCTGGGTTTAACTTAAGGCGCTGTTCCGTAATTT 672

Qy 171 AlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
Db 673 GCTGCTGCCCTTCTCTGGTATGGAGACACTCTCATCAGCGAAGCAGGAAGCATGGTTGAT 732
```

```
Qy 191 IleLysLeuGlu-----LysProGlnGlnThrValSerGluGlyGlyCys 206
Db 733 GTGAACCTTGAGGTCTGGCAATGCAAACTGCTCCAGTCTCAGGCTCAGGCTGGGGGATGC 792

Qy 207 SerCys 208
Db 793 AGTTGT 798

RESULT 15
US-10-425-114-27873
; Sequence 27873, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 27873
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4741-002-F12_FLI
US-10-425-114-27873

Alignment Scores:
Pred. No.: 1.25e-79 Length: 1308
Score: 705.50 Matches: 143
Percent Similarity: 72.46% Conservative: 28
Best Local Similarity: 60.59% Mismatches: 50
Query Match: 54.10% Indels: 15
DB: 17 Gaps: 3

US-09-441-857-12 (1-254) x US-10-425-114-27873 (1-1308)

Qy 11 LeuArgLysPheLysLeuValPheLeuGlyGlnSerValAlaLysThrSerLeuIle 30
Db 281 CTCGCCAAGTACAAAGCTCGTCTTCTGGGGAGCAGCTCCGCGGCAAGCAGCATCATC 340

Qy 31 ThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAlaIleIleGlyLeuAspPhe 50
Db 341 ACCCGCTTCATGATAGTAAAGTTCGACAACTTACCGAGGTACAAATTTGGTATTGATTTC 400

Qy 51 LeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLeuArgLeuTyrPheAspThrAla 70
Db 401 CTGTCAAAAGACAATGTACTTGAAGATAGAACTGTGAGACTCCAACCTCTGGATACCTGCT 460

Qy 71 GlyGlnGluArgLeuArgSerLeuIleProArgTyrIleArgAspSerAlaAlaVal 90
Db 461 GGTGCGGAAAGATTCAGGAGTTTAAATTCCAAGCTATATATTAGAGACTCTTCAGTTGCTGTC 520

Qy 91 ValValTyrAspIleThrAsnValAsnSerPheGlnGlnThrLysTrpIleAspAsp 110
Db 521 ATTGTATTCGATGTTCGAGCAGGAGCTCTTCTTAATATCTTAAGTGTGATAGAGAA 580

Qy 111 ValArgThrGluArgGlySerAspValIleIleThrLeuValGlyAsnArgThrAspLeu 130
Db 581 GTTCGCACTGAGAGGGGCGAGTGATTTATCTTGTCTTGTGGGAACAAACTGACCTT 640

Qy 131 AlaAspLysArgGlnValSerValGluGlyGlnArgLysAlaLysGlyLeuAsnVal 150
Db 641 GTTGCAAGAGCGCAAGTCTCAATAGAGGAAGGAGGCAAGGCAAGGACCTTTGGAGTC 700

Qy 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
Db 151 ThrPheIleGluThrArgAlaLysThrGlyTyrAsnValLysGlnLeuPheArgVal 170
```

```

Db      701 ATGTTTATTGAACCAAGCTGCTAAGCTGGGTTTAAACATTAAAGCGCTGTTCGTAAAAATT 760
Qy      171 AlaAlaLeuProGlyMetGluSerThrGlnAspGlySerArgGluAspMetSerAsp 190
      |||||
Db      761 GCTGTGCACCTTCCTGGATGGAGAGCTCTCATCAGCGAAACAGAGACATGGTTGAT 820
      |||||
Qy      191 IleLysLeuGlu-----LysProGlnGluGlnThrValSerGluGlyCys 206
      ::|||
Db      821 GTGAACCTTGAGGTCGCGCAATGCAAACTCGTCCAGTCTCAGGCTCAGGCTGGGGGATGC 880
      |||||
Qy      207 SerCysTyrSerProMetSerSerSerThrLeuProGlnLysProProTyrSerPheIle 226
      |||||
Db      881 AGTTGT-----TAGTTGCAGTCTCTGACACTTGCT----- 910
      |||||
Qy      227 AspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeuIleThr 242
      |||||
Db      911 ---TGCTCATGATCTTTGGGACTCTGAGCATCAGAGGCGTTTACT 955
      |||||

```

Search completed: April 25, 2005, 09:38:56  
 Job time : 6909 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: April 25, 2005, 02:37:36 ; Search time 3095 Seconds  
(without alignments)  
3123.853 Million cell up

```

Title: US-09-441-857-12
Perfect score: 1304
Sequence: 1 MSAGGDFGNPLRKFKLVELG.....NLFPSLITECNSSLLPVSWR 254

```

Scoring table: BLOSUM62

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```

Command line parameters:
-MODEL=frames_pzn.model -DEV=xlp
-Q/cn2_1/USPto.spool_p/US09441857/runat_22042005_155509_24759/app_query.fasta_1.391
-DB=EST -Qfmt=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MTRX=tbls06m2 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=est -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09441857 @CGN_1_5180 @runat_22042005_155509_24759 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRESH=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hlc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query #	SUMMARIES				Description
		Score	Match	Length	ID	
1	1304	100.0	3084	3	HS801695	AL136727 Homo sapi
2	1128	86.5	1928	3	AK084131	Mus muscu
3	1128	86.5	1955	3	AK051246	Mus muscu
4	1128	86.5	3404	3	AK083262	Mus muscu
5	1081	82.9	790	4	BI694279	603347724
6	1065	81.7	897	5	BQ422263	AGENCOURT
7	1016	77.9	744	6	CB320435	AGENCOURT
8	962	73.8	809	7	CN227000	RJB008D12
9	956.5	73.4	797	7	CF151490	AGENCOURT

10	953	73.1	792	6	CA316338	UI-M-TWO-	CA316338
11	944	72.2	628	6	CB565915	CB565915	AGENCY
12	941.5	72.2	870	6	CA471443	CA471443	AGENCY
13	936	71.8	615	6	CB067426	CB067426	1Q36F10.x
14	936	71.8	788	7	CK696452	CK696452	2F101-P00
15	934	71.6	696	5	BX952687	BX952687	BX852687
16	928	71.2	656	1	AI435940	AI435940	t180C04.x
17	927.5	71.1	803	5	BU379426	BU379426	603811211
18	925	70.9	866	1	AL875722	AL875722	AL875722
19	924	70.9	609	5	BQ632282	BQ632282	1125a01.x
20	923	70.6	838	7	CK315856	CK315856	SB02028A1
21	921	70.6	858	7	CF547798	CF547798	AGENCY
22	919	70.5	878	6	CD361397	CD361397	AGENCY
23	915	70.2	738	6	CA372914	CA372914	646911 NC
24	911	69.9	804	7	CK302212	CK302212	SB02015A1
25	909.5	69.7	759	5	BP709291	BP709291	BP709291
26	907	69.6	599	7	CV025778	CV025778	3550 Full
27	905.5	69.4	914	6	CA974052	CA974052	AGENCY
28	905	69.4	1864	3	CR692686	CR692686	Tetraodon
29	904	69.3	764	5	BU358285	BU358285	603476202
30	901.5	69.1	876	5	BQ732681	BQ732681	AGENCY
31	900	69.0	762	2	BE965054	BE965054	601658839
32	900	69.0	746	7	CO421669	CO421669	GGEZHT102
33	899	68.9	676	7	CV028279	CV028279	6756 Full
34	897.5	68.8	2563	3	AK035893	AK035893	Mus muscu
35	896	68.7	757	7	CK302742	CK302742	SB02016A2
36	894.5	68.6	715	4	BG016994	BG016994	df67Q09.x
37	894	68.6	871	5	BU381224	BU381224	603857076
38	893.5	68.5	833	7	CK352986	CK352986	AGENCY
39	891.5	68.4	854	5	BX734058	BX734058	BK734058
40	890	68.3	798	7	CF290541	CF290541	AGENCY
41	886	67.9	767	4	BI754357	BI754357	603028905
42	885	67.9	632	2	BF431157	BF431157	7608C05.x
43	882	67.6	792	7	CK081076	CK081076	EC2BBA18D
44	881.5	67.6	886	5	BU363505	BU363505	603785110
45	879	67.4	643	2	BB628557	BB628557	BB628557

```

/db_xref="RZPD:DKFZp566K144"
/db_xref="taxon:9606"
/clone="DKFZp566K144"
/tissue_type="kidney"
/clone_lib="566 (synonym: hfkd2). Vector pAMP1; host
xl-2blue; sites NotI + SalI"
/dev_stage="fetal"
/notes="RAB6C, member RAS oncogene family"
1..3084
/gene="DKFZp566K144"
456..1220
/gene="DKFZp566K144"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB6661.1"
/db_xref="GI:12052973"
/db_xref="GOA:Q9H0N0"
/db_xref="UniProt/Swiss-Prot:Q9H0N0"
/translation="MSAGSDFGNPLRKFVLCFQSVAKTSLITRPRYDSFDNTYQA
IIGIDFLSKTWLEDTGIGRLWDTAGQERLSLIPRYIRDSAAVVDYDITNVNSFQ
QITWIDVTRTGRSDVITLVGNRTDLADKROQSVSEGERKAKGLNVTFIETRAKTG
YNNQLFRVAAALPMESTQDGSREDMSDIKLEKPOQTIVSEGGSCSYSPMSSSTLP
QKPPYSFIDCSVNIGLNLFPSLITFCNSLLPVSWR"
ORIGIN
Alignment Scores:
Pred. No.: 5,696-135 Length: 3084
Score: 1304.00 Matches: 254
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 3 Gaps: 0

US-09-441-857-12 (1-254) x HSM801695 (1-3084)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
|||||
Db 456 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAATTCAGCTGGTGTCTCTGGGG 515
|||||

Qy 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40
|||||
Db 516 GAGCAAAAGCGTTGCAAGACATCTTTGATCACCAGATTCAGGTATGACAGTTTTGACAA 575
|||||

Qy 41 ThrTyrGlnAlaIleIleGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
|||||
Db 576 ACCTATCAGGCAATTAATTTGGCATTGACTTTTATCAAAACATATGATCTCGAGATGGA 635
|||||

Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
|||||
Db 636 ACAATCGGGCTTCGCTGGGATACGGGGGTGAGAACGCTCTCCGTAGCTCATTTCCC 695
|||||

Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
|||||
Db 696 AGGTACATCCGTGATTTCTGCTGCAGCTGTAGTTTACGATATCACAAATGTTAACTCA 755
|||||

Qy 101 PheGlnGlnThrThrLysThrIleAspAspValArgThrGluArgGlySerAspValIle 120
|||||
Db 756 TTCCAGCAAACTACAAAGTGGATTGATGATGTCAAGACAGAAAGGAGGATGTATTATC 815
|||||

Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
|||||
Db 816 ATCAGCTAGTAGGAATAGAACAGATCTTCTGACAGAGCAAGTGTCAAGTTGAGGAG 875
|||||

Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
|||||
Db 876 GGAGAGAGAAAGCCAAAGGGCTGAATGTTACGTTTATTGAAACTAGGCGCAAAACTGGA 935
|||||

Qy 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180
|||||
Db 936 TACAATGTAAGACGAGCTCTTTTCGACGTGTAGCAGCAGCTTTGCGCGGGAATCGAAGACA 995
|||||

Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
|||||

```

```

Db 996 CAGGCGGAAGCAGAGAGACATGAGTACATATAAACTGGAAAAGCCTCAGGAGCAACA 1055
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
|||||
Db 1056 GTACAGCAAGGGGTGTCTGCTACTCTCCCATGTCTTCAACCCCTTCTCCTCAGAAG 1115
Qy 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240
|||||
Db 1116 CCCCCTTACTCTTTTCATTCAGTCGAGTGTGAATATGGCTTGAACCTTTTCCCTTCATTA 1175
Qy 241 IleThrPheCysAsnSerSerLeuLeuProValSerTyrArg 254
|||||
Db 1176 ATAAGTTTTCGAATTCATTCATTCCTGCTGCTCTGCTGAGG 1217

RESULT 2
AK084131
LOCUS
DEFINITION
AK084131 1928 bp mRNA linear HTC 03-APR-2004
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
enriched library, clone:DI30097N11 product:RAB6, member RAS
oncogene family, full insert sequence.
ACCESSION
AK084131.1 GI:26350968
VERSION
HTC; CAP trapper.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2
20499374
MEDLINE
11042159
PUBMED
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
4
20530913
MEDLINE
11076861
PUBMED
5
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
6
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
7
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,

```

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sobabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

# TITLE JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.

## FEATURES

source

```
1..1928
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:DL130097N11"
/db_xref="taxon:10090"
/clone="DL130097N11"
/tissue_type="spinal ganglion"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
447..1073
/note="unnamed protein product: RAB6, member RAS oncogene family (MGDI|MG1:894313, GB|NM_024287, evidence: BLASTN, 99%, match=1919) putative"
/codon_start=1
/protein_id="BAC39121.1"
/db_xref="GI:26350969"
/translation="MSAGDGFNPLRKLVLPLGRQSVGKTSLTRMYVDFSDNYQA
TIGIDFLSKMYLEDRVLRLQWDVAGQERFSLIPSRDSTVAVVYDITNVNPFQ
QTTKWIDVTRGSDVIMLVGNKTDIADKRFQIEGRKAKELNMFYIETSAKAG
YNVKQLFRVAALPGMESTQDRSDREDMIDIKLEKPPQVPVNEGSC"
```

## CDS

237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254

1167 TTCCCTTCAGTAATAAGTATTGCAATTCATTCGTCGCTGTTTCGTGGAGG 1220

## RESULT 3

AK051246

LOCUS

DEFINITION

AK051246

1955 bp mRNA linear

Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length

enriched library, clone:DL130023A03 product:RAB6, member RAS

oncogene family, full insert sequence.

AK051246

AK051246.1 GI:26341819

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

PUBMED

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

687 AGCTACATTCGAGACTCCACTGTGGCAGTTCTGTTTATGATATCACAAATGTTAACTCA 746

101 PheGlnGlnThrThrLysThrLysThrLysThrLysThrLysThrLysThrLysThrLys 120

747 TTCGAGCAAACTACAAAATGGATTGATGTCAGAACGAGAGAGAGAGAGAGAGAGAG 806

121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140

807 ATCATGCTAGTAGGAATAAAACAGATCTTCTGATAGAGGCAAGTGTCAATTCAGGAG 866

141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheLeuThrArgAlaLysThrGly 160

867 GGAGAGAGAAAGCAAGAGAGCTGAATTTATTTATTTGAACAGTGTCAAAAGCAGGA 926

161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180

927 TACATGTAAAGCAGCTTTCCGGGTGTTCTGCGAGCTTTACCTGGAAATGGAAGCACA 986

181 GlnAspGlySerArgGluAspMetSerAspLysLeuGluLysProGlnGlnGlnThr 200

987 CAGCAGACAGAGAGAGACATGATTGACATAAACTGGAAAGCCCTCAGGAGCAACCA 1046

201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220

1047 GTCAATGAAGCGCGCTGTTCTCTGCTTAATCCCTCGGCATCTTTCAACCTCTCTGCAGAAG 1106

221 -----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236

1107 CTCAGCTGTTGGGCCCATCTCTTTCAATGACGTCAGTGTGAATATTGGCTTGAACCTT 1166

237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254

1167 TTCCCTTCAGTAATAAGTATTGCAATTCATTCGTCGCTGTTTCGTGGAGG 1220

## Alignment Scores:

Pred. No.: 1

Score: 1128.00

Percent Similarity: 89.53%

Best Local Similarity: 86.82%

Query Match: 86.50%

DB: 3

Length: 1928

Matches: 224

Conservative: 7

Mismatches: 23

Indels: 4

Gaps: 1

US-09-441-857-12 (1-254) x AK084131 (1-1928)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20

Db 447 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGAAATTCAGCTGTGTTCTGGGA 506

Qy 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40

Db 507 GAGCAGAGCGTTGGAAAGACGCTCTTGATCACCCTGATCATGATGATGATGATGATGAT 566

Qy 41 ThrTyrGlnAlaIleGlyLysPheLeuSerLysThrMetTyrLeuGluAspGly 60

Db 567 ACCATATCAGGCAACAATTCGCTGATCTTATCAAAACAAATGCTACTTGGAGGATAGA 626

Qy 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLysSerLeuIlePro 80

Db 627 ACCGTGCGGATTCATATTGGGACACAGCAGGTCAAGCGGTTCAGGAGCTGTGATCTCT 686

Qy 81 ArgTyrIleArgAspSerAlaAlaValValTyrAspIleThrAsnValAsnSer 100



```
99279253
10349636
2
Carninci, P., Shibata, Y., Hayateu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3404)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
Location/Qualifiers
1..3404
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C630031P05"
/db_xref="taxon:10090"
/clone="C630031P05"
/sex="male"
/tissue_type="hippocampus"

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
463..1089
/note="unnamed protein product; RAB6, member RAS oncogene
family (MGP|MGI:894313, GB|NM_024287, evidence: BLASTN,
99%, matches=1919)
putative"
/codon_start=1
/protein_id="BAC38834.1"
/db_xref="GI:26350389"
/translators="MSAGGDFGNLRKFKLVFLGQSVGKTSLLITRFMYDSFDNTYQA
TIGDPLSKTYLSDRVRLQNDTACGFRSLPSYIRDSRVAVVYVDITNVNSFO
QTTKWDVDRTERGSDVILMLVGNKTDLADKQSVIEGERKAKELNNVFIETSAKAG
YNVQLFRVRAALPGMESTQDRSREDMIDIKLEKPOBPQVNEGSCS"
3379..3384
polyA_signal
/note="putative"
3404
polyA_site
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 4,11e-115 Length: 3404
Score: 1128.00 Matches: 224
Percent Similarity: 89.53% Conservativeness: 7
Best Local Similarity: 86.82% Mismatches: 23
Query Match: 86.50% Indels: 4
DB: 3 Gaps: 1
US-09-441-857-12 (1-254) x AK083262 (1-3404)
QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
Db 463 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATCAAGCTGGTGTCTCTGGGA 522
QY 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAsp 40
Db 523 GAGCAGACGGTGGAAAGACGCTTGTATGATCACCCTGATCATGTATGACAGATTTTGAACAAC 582
QY 41 ThrTyrGlnAlaIleLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60
Db 593 ACCTATCAGGCAACATTTGGCATTTGACCTTTATCAAAAACAATGTACTTGGAGGATAGA 642
QY 61 ThrLeuGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuLeuPro 80
Db 643 ACCGTGCGATTGCAATTTATGGACACACGAGTCAAGACGCTTCAAGACGCTTGTATTCCT 702
QY 81 ArgTyrIleArgAspSerAlaAlaValValValValValValValValValValValVal 100
Db 703 AGCTACATTCGAGACTCCACTGTGGCAGTTGTTGTTATGATATCACAATGTTAACTCA 762
QY 101 PheGlnGlnThrThrLysThrIleAspAspValArgThrGluArgGlySerAspValIle 120
Db 763 TTCCAGCAAACTACAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822
QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 823 ATCATGCTAGTAGGAAATAAAACAGACTTCTGTGTAAGAGAGCAAGTGTCAATTCAGGAG 882
QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 883 GGAGAGAGGAAAGCCAAAGAGCTGAATGTTATTTTAAACCCAGTGCAGAAAGCAGAGA 942
QY 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaAlaLeuProGlyMetGluSerThr 180
Db 943 TACATGTTAAGCAGCTTTTCGGCGCTGTTGCTGAGCTTTTACCTGGAATGGAAGCACA 1002
QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 1003 CAGGACAGAGCAGAGAGACATGATTGACATATAAACTGGAAAGCCCTCAGGAGCAACCA 1062
QY 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
Db 1063 GTCAATGAAGCGCGCTGTTCTCTGCTTAATCCCTTCAACCTCTCTGAGAG 1122
```

```

Qy 221 -----ProPtyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236
Db 1123 CTCAGTCTTTGGCCCACTACTCTTCTTCACTGACGTGCAATATTGGCTTGAACCTT 1182
Qy 237 PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTyrArg 254
Db 1183 TTTCCCTTCAGTAATAACGATTGCAATTCATCTGCTGCCTGTTTCGTGGAGG 1236

RESULT 5
BI694279
LOCUS 603347724F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5375477 5',
DEFINITION mRNA sequence.
ACCESSION BI694279
VERSION BI694279.1 GI:15656908
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11955 row: p column: 06
High quality sequence stop: 788.
FEATURES
    source
    1..790
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="FVB/N-3"
    /db_xref="taxon:10090"
    /clone="IMAGE:5375477"
    /tissue_type="tumor, biopsy sample"
    /dev_stage="5 months"
    /lab_host="DH10B"
    /clone_lib="NCI_CGAP Mam2"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
    Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
    Library constructed by Life Technologies. Investigator
    providing samples: Gilbert Smith, NIH"

ORIGIN
Alignment Scores:
Pred No.: 9,33e-111 Length: 790
Score: 1081.00 Matches: 217
Percent Similarity: 88.24% Conservative: 8
Best Local Similarity: 85.10% Mismatches: 25
Query Match: 82.90% Indels: 5
DB: 4 Gaps: 1

US-09-441-857-12 (1-254) x BI694279 (1-790)

Qy 5 GlyAspPheGlyAsnProLeuArgLysPheLysLeuVal-PheLeuGlyGluInSerVa 24
Db 3 GGAGACTTCGGGAATCCGCTCAGGAAATTCAGCTGGTGGTTCCTGGGAGACGAGCGT 62
Qy 24 lAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGlnAl 44
Db 63 TGGAAAGACGCTCTTGATCACCCTGATTCATGTATGACAGCTTTTGCACACCTATCAGGC 122
Qy 44 aileileGlyIleAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGlyLe 64

```

```

Db 123 AACAAATTGGCATTCACCTTCTTATCAAAAACAATGCTACTTGGAGGATAGAACCGTCGCATT 182
Qy 64 uArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIleProArgTyrIleAr 84
Db 183 GCAATATTGGGACACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTGATTCTCTAGCTACATTGC 242
Qy 84 gApSerAlaAlaValValValTyrAspIleThrAsnValAsnSerPheGlnGlnTh 104
Db 243 AGACTCCACTGTGGCAGTTGTTGTTATGATATACAAAATGTTAACTCATTTCCAGCAAC 302
Qy 104 rThrLysTyrIleAspAspValArgThrGluArgGlySerAspValIleIleThrLeuVa 124
Db 303 TACAAAATGCTGATGATGTCAGAACAGAGAGGGAAGTGTGTCATCATCATGCTAGT 362
Qy 124 lGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGluGlyArgL 144
Db 363 AGGAAATAAACACAGATCTTCTGCTATAGAGGCAAGTGTCAATTGAGGAGGAGAGGAA 422
Qy 144 sAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGlyTyrAsnVal 164
Db 423 AGCCAAAGAGCTGAATGTTATGTTTATTTGAACCAAGTGCAGGAGCAGATACAAATGTAAA 482
Qy 164 sGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThrGlnAspGlySe 184
Db 483 GCAGCTTTTCGGCGTGTGCTGCAGCTTTACCTGGGAATGGAAGCACACAGGACAGAAG 542
Qy 184 rArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnGlnThrValSerGluGl 204
Db 543 CAGAGAGACATGATGATGACATAAAACTGGAAGACCTCAGGAGCAACAGCTCAATGAAGG 602
Qy 204 YGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys----- 220
Db 603 CGGCTGTCTCTGCTAAATCCCTCGGCATCTTTCAACCTCTCGAGAAGCTCACTGCTTT 662
Qy 221 -ProPtyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLe 240
Db 663 GGCCCCATACCTCTTTCATTCAGCTGCAGTGTGGAATATTGGCTTGAACCTTTTCCCTTCAGT 722
Qy 240 uIleThrPheCysAsnSerSerLeuLeuProValSerTyrArg 254
Db 723 AATAACGATTATGCAATTCATCAATTGCTGCTGTTTCTGTTGGAGG 765

RESULT 6
BI694279
LOCUS 7802212 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6042084
DEFINITION 5', mRNA sequence.
ACCESSION BI694279
VERSION BI694279.1 GI:21117590
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13280 row: O column: 13
High quality sequence stop: 595.
FEATURES
    source
    1..897
    /organism="Homo sapiens"
    /mol_type="mRNA"

```

CB320435  
ACCESSION  
CB320435.1 GI:2884670  
VERSION  
EST.  
KEYWORDS  
Mus musculus (house mouse)  
SOURCE

ORGANISM

100



```

Db      184 CCTAGCTACATTCGAGACTCCACTGTGGCAGTGTGTTTATGATATACAAATGTTAAAC 243
Qy      100 SerPheGlnGlnThrThrylsTrpIleAspAspValArgThrGluArgGlySerAspVal 119
      |||||
Db      244 TCATTCCAGCAAACTACAAAATGGATTGATGATGCAGAACAGAGAGAGGAAGTATGTC 303
Qy      120 IleIleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGlu 139
      |||||
Db      304 ATCATCATGCTAGTAGGAATAAACAAGATCTTGCTGATAGAGCAAGTGTCATTCAG 363
Qy      140 GluGlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThr 159
      |||||
Db      364 GAGGAGAGAGAGGAAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACCAAGTCAAAAGCA 423
Qy      160 GlyTyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSer 179
      |||||
Db      424 GGATACAATGTAAGACAGCTTTTCGGCGGTGTGCTGCAGCTTTTACCTGGGAATGGAAAGC 483
Qy      180 ThrGlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGln 199
      |||||
Db      484 ACACAGGACAGAGAGAGACATGATTGCATATAAAGCTGGAAAGCCCTCAGGAGCAA 543
Qy      200 ThrValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGln 219
      |||||
Db      544 CCAGTCAATGAAGGGCGGTGTTCTCTGCTAAATCCCCCTGGCATCTTTCAACCCCTCTCGCAG 603
Qy      220 lys-----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsn 235
      |||||
Db      604 AAACCTACTGCTTTGGCCCCCCTACTCTTTGCTGATGACTGCAGTGTGAATATTGGCTTGAAC 663
Qy      236 Leu-PheProSerLeuIleThrPheCysAsnSerSerLeuLeuProValSerTrpArg 254
      |||||
Db      664 CTTTTCCTTCAGTAATAAGATGATTGCAATTCATCATTTGCTGCTGTTTCGTGGAGG 721

```

## RESULT 8

```

LOCUS      CN227000                      809 bp    mRNA    linear    EST 09-APR-2004
DEFINITION RJB008D12.ab1 RUTestis Gallus gallus cdna 5', mRNA sequence.
ACCESSION  CN227000
VERSION     CN227000.1  GI:46330744
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 809)
            Savolainen,P., Fitzsimmons,C.J., Arvestad,L., Andersson,L. and
            Lundeberg,J.
            EST analysis of brain and testis cdna libraries from White leghorn
            and Red Jungle Fowl
            Unpublished (2004)
JOURNAL    Contact: Peter Savolainen
COMMENT    Department of Biotechnology
            Royal Institute of Technology, KTH
            SE-106 91 Stockholm, SWEDEN
            Tel: +46 (0)8 5537 8481
            Fax: +46 (0)8 5537 8335
            Email: Peter.savolainen@biotech.kth.se
            Seq primer: M13 reverse primer.
            Location/Qualifiers
                1..809
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Red Junglefowl"
                /db_xref="taxon:9031"
                /sex="male"
                /lab_host="ElectroMAX DH10B (Invitrogen)"
                /clone_lib="RUTestis"
                /note="Organ: testis; Vector: pSPORT-1; Site 1: Hind III;
                Site 2: EcoRI; The cdna libraries were created with the
                Superscript Plasmid System (Invitrogen)."
```

## FEATURES

```

source     1..809
            CF151490                      797 bp    mRNA    linear    EST 25-JUL-2003
            AGENCOURT 14907987 NICHD_XGC Emb8 Xenopus tropicalis cdna clone
            IMAGE:6983585 5', mRNA sequence.
            CF151490
            VERSION     CF151490.1  GI:33247998
            KEYWORDS   EST.
            XENOPUS tropicalis (western clawed frog)
            ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.
            REFERENCE  1 (bases 1 to 797)

```

```

Alignment Scores:      2.34e-97      Length:      809
Score:                 962.00      Matches:     201
Percent Similarity:    87.03%      Conservative: 7
Best Local Similarity: 84.10%      Mismatches:  28
Query Match:           73.77%      Indels:       6
DB:                    7           Gaps:         1

```

US-09-441-857-12 (1-254) x CN227000 (1-809)

```

Qy      1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20
      |||||
Db      92 ATGTCGGCGGGGGGACTTCGGCAACCCGCTCGGAAATTCAGCTGGTGTCTCTGGGC 151
Qy      21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsn 40
      |||||
Db      152 GAGCAGAGCGTGGGGAAGACTTCCTTGATCACCAGGTTTCATGATGACAGCTTTGACAAC 211
Qy      41 ThrTyrGlnAlaIleIleGlyIleAspPheIleuSerLysThrMetTyrLeuGluAspGly 60
      |||||
Db      212 ACCTACCAGGCACAAATTGGCATTTGACTTCTTATCAAAAACCATGTATTTGGAGGATCGA 271
Qy      61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80
      |||||
Db      272 ACAATCAGGTTGGAGCTGTGGATACTCGGGTCAGGAACGTTTCCGTAGCTCATTTCCC 331
Qy      81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100
      |||||
Db      332 AGTTACATCGGTGACTCTGCTGCTGTAGTAGTTTACGATATCACAATGTCAACTCG 391
Qy      101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120
      |||||
Db      392 TTCAGCAAAACAAATAATGGATCGACGATGTGAGAACCGGACGGCGATGTGCATC 451
Qy      121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
      |||||
Db      452 ATTATGCTGGTGGGAATAATAAACAGATCTAGCAGATAAGAGCAAGTGTCTATTGAGAA 511
Qy      141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
      |||||
Db      512 GGAGAAAGAAAGCCAAAGAGCTGAATGTAATGTTTCATCGAAACTAGTGCAGAAAGCAGGA 571
Qy      161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180
      |||||
Db      572 TACAATGTAAACACAGCTTTTCCGACGCTGGCAGCTGCCTTGCCTTGGATGGAAGACACA 631
Qy      181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
      |||||
Db      632 CAAGACAAAAGTAGAGACATGATTGACATCAAACTGGAAGAGCTCAAGAGCACCCCT 691
Qy      201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220
      |||||
Db      692 GTCAGTGAAGAGGCTGCTCCTGCTAATACACAGCTGGCTTCTGACTTCCT-CCAGAACAT 750
Qy      221 -----ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeu 236
      |||||
Db      751 CACTGCTTTCCCTCCCTT-ACTCTTATTGACTGCAGGGG-GAAATTTGGCTTTGAACCTT 805

```

## RESULT 9

```

LOCUS      CF151490
DEFINITION AGENCOURT 14907987 NICHD_XGC Emb8 Xenopus tropicalis cdna clone
ACCESSION  CF151490
VERSION     CF151490.1  GI:33247998
KEYWORDS   EST.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
            Xenopodinae; Xenopus; Silurana.
            REFERENCE  1 (bases 1 to 797)

```



# AUTHORS TITLE JOURNAL COMMENT

NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics / NIH  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Robert M. Grainger  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM14648 row: d column: 16  
High quality sequence stop: 708.  
Location/Qualifiers  
1. .797  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:5983585"  
/tissue\_type="cadpole"  
/dev\_stage="embryo, stages 40-45"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NICHG XGC Emb8"  
/note="Vector: pCIV-SFORS.1; Site.1: NotI; Site.2: EcoRV;  
Cloned unidirectionally. Primer: Oligo dt. Average insert  
size 2.1 kb. Constructed by Invitrogen. Note: This is a  
Xenopus Gene Collection (XGC) library."

## FEATURES source

### ORIGIN

Alignment Scores:  
Pred. No.: 9,52e-97 Length: 797  
Score: 956.50 Matches: 193  
Percent Similarity: 80.08% Conservative: 8  
Best Local Similarity: 76.89% Mismatches: 29  
Query Match: 73.35% Indels: 21  
DB: 7 Gaps: 1

US-09-441-857-12 (1-254) x CFI51490 (1-797)

Qy 1 MetSerAlaGlyValAspPheGlyAenProLeuArgLysPheLysLeuValPheLeuGly 20  
Db 75 ATGTCGGCGGAGGACTTCGGGAACCGCTGAGGAATTCAACTGTCTCTCTGGGG 134  
Qy 21 GluGlnSerValAlaLysThrSerLeulleThrArgPheArgTyrAspSerPheAspAen 40  
Db 135 GAGCAGAGCGTTGGGAAGACTTCTTTAATAACAGGTTTCATGTATGACAGCTTTGACAAAC 194  
Qy 41 ThrTyrGlnAlaIleIleGlyLeaAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 195 ACATACCAGCAACAACTTGAATTCATCTTATCCAAAACAAATGTACCTAGAGGATCGA 254  
Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeullePro 80  
Db 255 ACAGTCGGGTTCGAATTTGGGACACAGCAGGTTCAGGACGGTTTCAGAGATTTCCTCT 314  
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAenValAenSer 100  
Db 315 AGCTACATTCGGGACTCCACTGTGGCAGTTGTTTATTTGATATCACAATGTAAACTCA 374  
Qy 101 PheGlnGlnThrThrylserPheAspValArgThrGluArgLysSerAspValIle 120  
Db 375 TTCCAGCAACACCCAGAAATGGATTCGATGACGTTCGGAACAGAGCGGGGGGACGCGTATC 434  
Qy 121 IleThrLeuValGlyAenArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
Db 435 ATTATGTTAGTGGGAATTAACCCGACCTGGCAGATAAAGACNAGTGTCCATTGAGGAA 494  
Qy 141 GlyGluArgLysAlaLysGlyLeuAenValThrPheIleGluThrArgAlaLysThrGly 160

Db 495 GGAGAGCGGAAGCAAGGAGCTGAACGTAATGTTTATTGAAACGAGTGCAAGGCGCGC 554  
Qy 161 TyrAenValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180  
Db 555 TACATGTAAAGCAGCTCTTCGACCGCTGGCGCTGCTGCTGGATGGAAGCTCA 614  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluThr 200  
Db 615 CAGGACAAGAGCAGAGAAGACATGATTGATATCAACTGGAAAAGCCTCCAGAACAGCCA 674  
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerSerThrLeuProGlnLys 220  
Db 675 GTGTCGGAAGGGGCTGCTCTGTTATGCCCATCCCTCCCGGAGCTCCCCCATGAT 734  
Qy 221 ProProTyrSerPheIleAspCysSerValAenIleGlyLeuAenLeuPheProSerLeu 240  
Db 735 CTTCCC----- 740  
Qy 241 lleThrPheCysAenSerSerLeuLeuProVal 251  
Db 741 -----TCCGGGACATCACTGCTCTCTGTC 764

RESULT 10  
CA316338  
LOCUS  
DEFINITION  
IMAGE:6812737 5', mRNA sequence.  
ACCESSION  
CA316338.1 GI:24534462  
VERSION  
EST.  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
REFERENCE  
1 (bases 1 to 792)  
AUTHORS  
NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

## FEATURES source

Location/Qualifiers  
1. .792  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6812737"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_FW0"  
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Ronald, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is ACCGAGACAG. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

```

ORIGIN
Alignment Scores:
Pred. No.:      2,34e-96      Length:      792
Score:          953.00      Matches:     194
Percent Similarity: 86.90%      Conservative: 5
Best Local Similarity: 84.72%      Mismatches:  22
Query Match:    73.08%      Indels:      8
DB:             6          Gaps:       2

US-09-441-857-12 (1-254) x CA316338 (1-792)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLeuValPheLeuGly 20
Db 65 ATGTCGCGGGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGA 124

Qy 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40
Db 125 GAGCAGAGCGGTGGAAAGACGCTTGTATCATCCCGGATTCATGTATGACAGTTTGGACAAAC 184

Qy 41 ThrTyrGlnAlaLysLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 60
Db 185 ACCTATCAGGCAACATTTGGCATTCGCTTCTTATCAAAACATGATGATGAGGATAGA 244

Qy 61 ThrLeuGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuLeuPro 80
Db 245 ACCGTGCGATTGCAATTTATGGGACACAGCAGGTCAAGAGCGGTTTCAGGAGCTTGATTCCT 304

Qy 81 ArgTyrLeuArgAspSerAlaAlaValValValValValValValValValValValVal 100
Db 305 AGCTACATTCGAGACTCCACTGTGCGACGTGTGTTTATGATATACAAAATGTTAACTCA 364

Qy 101 PheGlnGlnThrThrLysTrpIleAspValArgThrGluArgGlySerAspValIle 120
Db 365 TTCAGCAACACTACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATC 424

Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db 425 ATCATGCTAGTAGGAATAAAACAGATCTTGCTGATAAGAGCAAGTGTCTATTTGAGGAG 484

Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db 485 GGAGAGAGAGAAAGCCAAAGAGCTGAATGTATGTTTATTGAAACCCAGTGCAAAAGCAGGA 544

Qy 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaAlaLeuProGlyMetGluSerThr 180
Db 545 TACATGTATAGCAGCTTTTCGGGCTGTGTCTGCGAGCTTTACCTGGGAATGGAAAGCACA 604

Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThr 200
Db 605 CAGGACAGAAGCAGAGAAGACATGATTGACATAAACTGGAAAGCCCTCAGGAGCAACCA 664

Qy 201 ValSerGluGlyGlyCysSerCysTyrSerPro-----MetSerSerSer 215
Db 665 GTCAATGAAGCGCGGTGTCTCTGCTTAATCCCTCGCATCTTCAACCCCTCTCGCAGACTCA 724

Qy 216 ThrLeuProGlnLysProProTyrSer 224
Db 725 CTGCTT-----GGCCCATACTCT 742

```

```

RESULT 11
LOCUS      CB565915
DEFINITION AGENCOURT 12691469 NIH MGC 146 Homo sapiens cDNA clone
IMAGE:6519185 5', mRNA sequence.
ACCESSION CB565915
VERSION    CB565915.1 GI:29485445
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 628)  
NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: csapbs-remail.nih.gov

Tissue Procurement: Guthrie cDNA Resource Center

cDNA Library Preparation: Guthrie cDNA Resource Center

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: IRBF001 row: d column: 05

High quality sequence stop: 628.

FEATURES

Source

1..628

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mrna"

/db\_xref="taxon:9606"

/clone="IMAGE:6519185"

/tissue\_type="mixed"

/lab\_host="DH10B (Ti-phage-resistant)"

/clone\_lib="NIH\_MGC\_146"

/note="vector: pCDNA3.1; Site 1: multiple; Site 2:

multiple; ORF's were PCR-amplified (from IMAGE clones or

from commercially available cDNA libraries) and cloned by

the Guthrie cDNA Resource Center (www.guthrie.org/cDNA)

into pCDNA3.1. For specific information on cloning sites

(which vary by clone), please refer to the Guthrie

website, using the Guthrie ID given in the file

ftp://image.llnl.gov/image/rearrayed\_plates/IRBF.presv.dat

a. Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1.71e-95 Length: 628

Score: 944.00 Matches: 188

Percent Similarity: 91.83% Conservative: 3

Best Local Similarity: 90.38% Mismatches: 17

Query Match: 72.39% Indels: 0

DB: 6 Gaps: 0

US-09-441-857-12 (1-254) x CB565915 (1-628)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLeuValPheLeuGly 20

Db 1 ATGTCCACGTGCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCTGGGG 60

Qy 21 GluGlnSerValAlaLysThrSerLeuLeuThrArgPheArgTyrAspSerPheAspAsn 40

Db 61 GAGCAAGCGTTGGAAAGACATCTTTCATCACCAGATTCATGTATGACAGTTTTGACAAAC 120

Qy 41 ThrTyrGlnAlaLysLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60

Db 121 ACCATCAGGCAACAATTTGGCATTTGACTTTTATCAAAAAACTATGTACTTGGAGGATCGA 180

Qy 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80

Db 181 ACAATCAGGCTTCAGCTGTGGGATACCTCGGGTCAGGAACGTTCCGTAGCCTCATTTCCC 240

Qy 81 ArgTyrIleArgAspSerAlaAlaValValValValValValValValValValValVal 100

Db 241 AGTTACATCCGTGATCTGCTGCGAGCTGTAGTAGTTTACGATATACAAAATGTTAACTCA 300

Qy 101 PheGlnGlnThrThrLysThrIleAspAspValArgThrGluArgGlySerAspValIle 120

Db 301 TTCAGCAAACTCAAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATC 360

Qy 121 IleThrLeuValGlyLysAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140

Db 361 ATCATGCTAGTAGGAATAAACAAGATCTTGGTACAGAGCAAGTGTCAATTGAGGAG 420  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 421 GGAGAGAGAGAAAGCCAAAGAGCTGAATGTTATTTGAACTAGTGCCTCAAGCTGGA 480  
Qy 161 TyrAsnValLysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThr 180  
Db 481 TACAATGTAAGCAGCTCTTTCGACGTGTAGCAGCAGCTTTCGCGGAATGGAAGCACA 540  
Qy 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluThr 200  
Db 541 CAGGACAGAGAGAGAGAAATGATGATGACATAAACTGGAAGAGCCTCAGGAGCAACA 600  
Qy 201 ValSerGluGlyGlyCysSerCys 208  
Db 601 GTCAGTGAAGGAGGCTGTTCTCTGC 624

## RESULT 12

CA471443 870 bp mRNA linear EST 12-NOV-2002  
LOCUS AGENCOURT 10698065 NCI CGAP ZK1d1 Danio rerio cDNA clone  
DEFINITION IMAGE:6791338 5', mRNA sequence.

CA471443  
VERSION CA471443.1 GI:24927795

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.

## REFERENCE

NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## AUTHORS

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue procurement: Leonard I. Zon, M.D.

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM14294 row: n column: 09

High quality sequence stop: 721.

## FEATURES

source

1..870

Location/Qualifiers

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="IMAGE:6791338"

/lab\_host="DH10B (71-resistant)"

/clone\_lib="NCI CGAP ZK1d1"

/note="Organ: kidney; Vector: pCMV-SPORT6.1; Site 1:  
EcoRV; Site 2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.8 kb. Constructed by J.  
Wang (Research Genetics, Invitrogen Corp) from tissue  
donated by L. Zon (Harvard University). Note: this is a  
NCI CGAP Library."

## ORIGIN

Alignment Scores:

Pred. No.: 5,27e-95 Length: 870  
Score: 941.50 Matches: 193  
Percent Similarity: 79.84% Conservative: 13  
Best Local Similarity: 74.81% Mismatches: 35  
Query Match: 72.20% Indels: 17  
DB: 6 Gaps: 2

US-09-441-857-12 (1-254) x CA471443 (1-870)

Qy 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20

Db 87 ATGTCGACGAGGAGAGATTTCGGCAACCCGCTCCGCAAAATTCAAGCTAGTTTTCTGGGA 146  
Qy 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAn 40  
Db 147 GAGCAGACGGTGGGAAGACGTCGTGTGATTACCAATTTATGTACAGCAGTTTCCGATAC 206  
Qy 41 ThrTyrGlnAlaIleIleGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGly 60  
Db 207 ACCTACCAAGGCCACCATAGGAATTGTTCTTGCGAAACCATGATGATCTTGGAGCAGA 266  
Qy 61 ThrIleGlyLeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIlePro 80  
Db 267 ACAGTGCAGCTTCGAGCTTTGGGACACAGCAGCAGGAGCGTTTCCGAGCTGTGATCCC 326  
Qy 81 ArgTyrIleArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSer 100  
Db 327 AGCTACATCCGGGACTCGACGGTGTCTGTGTATGATCATCACAATGTCAACTCA 386  
Qy 101 PheGlnGlnThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIle 120  
Db 387 TTCCAGCAGACACCACCAATGGATCGATCGATCGGACCGAAAGAGGAAGTGTATCATC 446  
Qy 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGly 140  
Db 447 ATCATGCTAGTGGGAAACAAACACAGATCTTCGACACAAAAGCAAGTGTCTATCGAGAA 506  
Qy 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
Db 507 GGAGAAAGGAAGCCAAAGAACTGAATGTAATGTTTATGAAACTAGTGTCTAAAGCCGA 566  
Qy 161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180  
Db 567 TACAACGTCAGCAGCTTTTCCGCGGTGTAGCAGCAGCCCTTCTGTGTATGGAGACACA 626  
Qy 181 GlnAspGlySerArgGluAspMetSerAspLysLeuGluLysProGlnGluThr 200  
Db 627 CAAGACAAGACAGCAGAGCAGATCATGCACATAAAATTGGAAAGCCACCAGACGCG 686  
Qy 201 ValSerGluGlyGlyCysSerCysTyrSerProMetSerSerThrLeuProGlnLys 220  
Db 687 GTCAGGAAGCGCGCTGTTTCATGCTGAGGC-----SerSerLeuLeuProValSerTrp 716  
Qy 221 ProProTyrSerPheIleAspCysSerValAsnIleGlyLeuAsnLeuPheProSerLeu 240  
Db 717 -----TTCAGTCCGCTGTTCTGTGCGCGTCTCTTACTGTGCTCTTTTCAGACGCTCT 770  
Qy 241 IleThrPheCysAsn-----SerSerLeuLeuProValSerTrp 253  
Db 771 TATTCGCATCTCACAGCATCTCTCTGGATAAATCTGCAGAAGTCCGACCATGCG 824

## RESULT 13

CB067426

LOCUS

CB067426

DEFINITION

ig36f10.x1 HR85 islet Homo sapiens cDNA clone IMAGE: 3', similar to

SW:RAB6 HUMAN P20340 RAS-RELATED PROTEIN RAB-6. [1] ; mRNA

sequence.

ACCESSION

CB067426

VERSION

CB067426.1 GI:27811946

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 615)

Melton, D., Brown, J.,

Kenty, G., Permutt, A., Lee, C.,

Kaestner, K.,

Lemishka, I., Searce, M.,

Brestelli, J., Gradwohl, G.,

Clifton, S.,

Hillier, L., Marra, M.,

Pape, D., Wylie, T.,

Martin, J., Blistain, A.,

Schmitt, A., Theising, B.,

Ritter, E., Ronko, I.,

Bennett, J.,

Cardenas, M., Gibbons, M.,

McCann, R., Cole, R.,

Teagareishvili, R.,

Williams, T., Jackson, Y.,

and Bowers, Y.

Endocrine Pancreas Consortium

Unpublished (2000)

TITLE

JOURNAL

## COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohph.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Gibco  
High quality sequence stop: 444.

## FEATURES

source

1. .615  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_lib="HR85 islet"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
NotI; Site\_2: XhoI; cDNA made by oligo-dT priming.  
Size selected on agarose gel. Average insert size ~1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

## ORIGIN

Alignment Scores:  
Pred. No.: 1,32e-94 Length: 615  
Score: 936.00 Matches: 186  
Percent Similarity: 92.20% Conservative: 3  
Best Local Similarity: 90.73% Mismatches: 16  
Query Match: 71.78% Indels: 0  
DB: 6 Gaps: 0

US-09-441-857-12 (1-254) x CB067426 (1-615)

QY 4 GlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGlyGluGlnSer 23  
DB 1 GCGGAGACTTCGGGAATCCGCTGAGGAAATTCAGCTGGTGTCTCGGGGAGCAAAGC 60  
QY 24 ValAlaLysThrSerLeuIleThrArgPheArgTyrAspSerPheAspAsnThrTyrGln 43  
DB 61 GTTGAAGAAGACATCTTTGATCACCAGATTGATGATGATGATGATGATGATGATGATGAT 120  
QY 44 AlaIleLeuGlyLeuAspPheLeuSerLysThrMetTyrLeuGluAspGlyThrIleGly 63  
DB 121 GCAACAATGGCATTTGATTTTATCAAAAATACTGTTACTTGGAGATCGAACATCAGG 180  
QY 64 LeuArgLeuTyrAspThrAlaGlyGlnGluArgLeuArgSerLeuIleProArgTyrIle 83  
DB 181 CTTACGCTGTGGGATACCTCGGGTCAGGAACGTTTCCTAGCTCATTTCCAGTTACATC 240  
QY 84 ArgAspSerAlaAlaValValValTyrAspIleThrAsnValAsnSerPheGlnGln 103  
DB 241 CGTGATTTCTGTGACGCTGTAGTGTATACATATACAAATGTTAACTCATTCACGCAA 300  
QY 104 ThrThrLysTrpIleAspAspValArgThrGluArgGlySerAspValIleIleThrLeu 123  
DB 301 ACTACAAAGTGGATTGATGATGTCAGACAGAGAGGAAGTGTATGTTATCATCATGCTA 360  
QY 124 ValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlyGluArg 143  
DB 361 GTAGGAATAAAACAGATCTTGCTGACAAAGAGCAAGTGTCAATTTGAGGGAGAGAGG 420

QY 144 LysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGlyTyrAsnVal 163  
DB 421 AAAGCCAAAGAGCTGAATGTTATGTTTATTGAAACTAGTCAAAAGCTGGGATACAAATGA 480  
QY 164 LysGlnLeuPheArgValAlaAlaLeuProGlyMetGluSerThrGlnAspGly 183  
DB 481 AAGCAGCTCTTCCGCGTGTAGCAGCAGCTTTCGGGGAATGGAAAGCACACAGGACAGA 540  
QY 184 SerArgGluAspMetSerAspIleLysLeuGluLysProGlnGluGlnThrValSerGlu 203  
DB 541 AGCAGAGAAGATATGATTGACATAAACTGGAAGAGCTCAGGAGCAACCACTCAGTGAA 600  
QY 204 GlyGlyCysSerCys 208  
DB 601 AGAGGCTGTTCTCTGC 615

## RESULT 14

CK696452

LOCUS

DEFINITION

IMAGE:7161798.5', mRNA sequence.

ACCESSION

CK696452

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE

1 (bases 1 to 788)

AUTHORS

Wei, C., Mathavan, S., Thoreau, H., Lim, L., Lee, C. and Ruan, Y.

TITLE

Genome Institute of Singapore, Zebrafish Gene Collection

JOURNAL

Unpublished (2004)

COMMENT

Contact: Ruan Y

Cloning and Sequencing

Genome Institute of Singapore

60 Biopolis Street, #02-01, Genome, Singapore 138672

Tel: +65 6478 8073

Fax: +65 6478 9059

Email: ruanyj@gis.a-star.edu.sg

GIS Clone ID: ZF101-P00073-BR2\_F04

PCR Primers

FORWARD: M13

BACKWARD: M13

Plate: ZF101-P00073-BR2 row: F column: 04

Seq primer: CCGCATAACTGTATAGCA

High quality sequence stop: 788.

Location/Qualifiers

1..788

/organism="Danio rerio"

/mol\_type="mRNA"

/strain="Singapore local strain"

/db\_xref="taxon:7955"

/clone="IMAGE:7161798"

/tissue\_type="Embryo"

/dev\_stage="7 Different embryonic Stages (From just

fertilized Embryos to 72 hours just hatched baby fish)"

/lab\_host="DH10B"

/clone\_lib="GISZF001 ra"

/note="Vector: pDNR-LIB; Site 1: Sfi A (GGCATTACGGCC);

Site 2: Sfi B (GGCGAGGGGCC); Priming method: Sfi-(dT)30

Primed; Priming sequence:

5.ATTCAGAGCGAGCGCGGCACATG(T)30VN; Directionally

cloned, 5' cloning site: Sfi A site GGCATTACGGCC; 5'

linker/adaptor sequence: 5.AGACAGTGTATCAACGAGAGTGGCC;

3' cloning site: Sfi B site GGCCGAGGGGCC; 3'

linker/adaptor sequence: same as the priming sequence;

Average insert size: 2kb; For PCR insert analysis: Use

M13 Forward and reverse primers; Library Amplified;

Recombinants (inserts): 98%; Library complexity: 5x10<sup>6</sup>;

Full-length construction (method): SMART, a Clontech

method The pooled tissue RNA was collected and used to

construct full length enriched cDNA library and also

served as template to synthesize complex first strand cDNA probe. Two high density colony arrays were made from over 110K cDNA clones and hybridized with the probes. Low intensity clones were selected as they represented rare expressed clones. The hybridization intensities for all clones span from 0 to 1.8 million counts and the low abundant class ranged from 0 to 13,000."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,898-94 Length: 788  
 Score: 936.00 Matches: 185  
 Percent Similarity: 90.87% Conservative: 4  
 Best Local Similarity: 88.94% Mismatches: 19  
 Query Match: 71.78% Indels: 0  
 DB: 7 Gaps: 0

US-09-441-857-12 (1-254) x CK696452 (1-788)

QY 1 MetSerAlaGlyGlyAspPheGlyAsnProLeuArgLysPheLysLeuValPheLeuGly 20  
 |||||  
 Db 142 ATGTCGACGAGGAGATTTCGCAACCGCTCCGCAATTCAAGTAGTTTCTGGGA 201  
 |||||

QY 21 GluGlnSerValAlaLysThrSerLeuIleThrArgPheArgTyAspSerPheAspAsn 40  
 |||||  
 Db 202 GAGCAGAGCGTGGAAAGACGTCGTTGATTACAGATTATGACACAGATTTCGATAAC 261  
 |||||

QY 41 ThrTyGlnAlaIleThrLeuGlyLeuAspPheLeuSerLysThrMetTyLeuGluAspGly 60  
 |||||  
 Db 262 ACCTACCAAGCCACCATAGGAATTGATTTCTGTGAAACCATGATCTTGGAGACAGA 321  
 |||||

QY 61 ThrIleGlyLeuArgLeuTrpAspThrAlaGlyGlnGluArgLeuSerLeuIlePro 80  
 |||||  
 Db 322 ACAGTGGCGCTCAGCTTTGGACACAGCAGGACGCTTTCGCGAGTCTGATTCC 381  
 |||||

QY 81 ArgTyIleArgAspSerAlaAlaValValValValTyAspIleThrAsnValAsnSer 100  
 |||||  
 Db 382 AGCTACATCCGGGACTCGACGCTGCTGTGTGTGATCTATGACATCACAATGTCAACTCA 441  
 |||||

QY 101 PheGlnGlnThrThyLysTrpIleAspValArgThrGluArgLysSerAspValIle 120  
 |||||  
 Db 442 TTCACGACAGACCACCAATGGATGATGATGATGATGATGATGATGATGATGATGATC 501  
 |||||

QY 121 IleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140  
 |||||  
 Db 502 ATCATCTAGTGGGAAACAAACAGATCTTGAGATAAAGGCAAGTGTCTATCGAGGNA 561  
 |||||

QY 141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160  
 |||||  
 Db 562 CGAGAAAGGAAAGCCAAAGAACTGAATGTAATGTTATTGAAACTAGTGTAAAGCCGA 621  
 |||||

QY 161 TyrAsnValLysGlnLeuPheArgValAlaAlaAlaLeuProGlyMetGluSerThr 180  
 |||||  
 Db 622 TACAACGTCAACAGCTTTTCGCGCGTGTGAGCAGCAGCCCTTCCTGGTATGAGAGACA 681  
 |||||

QY 181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200  
 |||||  
 Db 682 CAAGACAGAGCAGAGAGACATGATCGACATATAAATTTGAAAAGCCGCCAGAGCCA 741  
 |||||

QY 201 ValSerGluGlyGlyCysSerCys 208  
 |||||  
 Db 742 GTCAGCGAAGCGCGCTGTTCTTTCG 765  
 |||||

## RESULT 15

BX852687  
 LOCUS BX852687 NICHD XGC Emb4 Xenopus laevis mRNA linear EST 11-DEC-2003  
 DEFINITION ; IMAGE:5541856 5', mRNA sequence.  
 ACCESSION BX852687  
 VERSION BX852687.1 GI:39740496  
 KEYWORDS  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus.

## REFERENCE

1 (bases 1 to 896)

## AUTHORS

Heil, O., Neubert, P., Peters, M., Radelof, U., Schneider, D.,

## TITLE

Schroth, A., Korn, B. and Landgrebe, J.

## JOURNAL

Xenopus laevis UniGene Set 1 (RZPDLIB No.988)

## COMMENT

Unpublished (2003)

Contact: Ina Rolfes

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; IMAGp998D1712240.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection (amp- resistant) (RZPDLIB

No.998) http://www.rzpd.de/cgi-

bin/products/showLib.pl.cgi/response?libNo=998 RZPDLIB; Xenopus

laevis UniGene Set 1 (RZPDLIB No.988) http://www.rzpd.de/cgi-

bin/products/showLib.pl.cgi/response?libNo=988 Contact: Ina Rolfes

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

www.rzpd.de

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

SP6, 5' ATTTAGGTGACACTATAG 3'

Location/Qualifiers

1..696

/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

/clones="IMAGp998D1712240 ; IMAGE:5541856"

/dev\_stage="embryo, stage 31-32"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NICHD XGC Emb4"

/note="Organ: whole embryo; Vector: pCMV-SPORT6; Site: 1;

NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo

dt. Average insert size 2.1 kb. Constructed by Life

Technologies. Note: This is a Xenopus Gene Collection

(XGC) library."

```
Db      344 TTCCAGCAAAACACAAATGGATCGATGATGTCCGAACAGAGCGGGGAGTGATGTGATC 403
Qy      121 lleThrLeuValGlyAsnArgThrAspLeuAlaAspLysArgGlnValSerValGluGlu 140
Db      404 ATTATGCTTGTGGGAAATAAAACAGATCTTGGGATAAAAGGCAAGTGTCATTGAGGAA 463
Qy      141 GlyGluArgLysAlaLysGlyLeuAsnValThrPheIleGluThrArgAlaLysThrGly 160
Db      464 GGGGAGCGGAAAGCCAGAGCTGAACGTAATGTTATTGAAACCAGTGCAAGGCTGGC 523
Qy      161 TyrAsnValLysGlnLeuPheArgArgValAlaAlaLeuProGlyMetGluSerThr 180
Db      524 TACAATGTAAAGCAGCTCTTCCGACGTGTTGCAGCGCGCTTGCCCTGGATGGAAAGCTCA 583
Qy      181 GlnAspGlySerArgGluAspMetSerAspIleLysLeuGluLysProGlnGlnThr 200
Db      584 CAGGATAAGAGCAGAGAGACATGATTGATATCAAACTGGAAAGCCCTCCAGAACAGCCA 643
Qy      201 ValSerGluGlyGlyCysSerCys 208
Db      644 GTGTCGGAAGGAGGCTGCTCTGT 667
```

Search completed: April 25, 2005, 05:40:55  
Job time : 3110 secs